

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:34 ; Search time 5799.58 Seconds
(without alignments)
10679.608 Million cell updates/sec

Title: US-09-930-440B-1
Perfect score: 1429
Sequence: 1 atggccttcocaaagaagaa.....ctttgaatttcatttcgatg 1429

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank:

- 1: gb_ba.*
- 2: gb_btg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_in.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by charge to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1429	100.0	1429	6	BD270236	Human gly
2	1332.4	93.2	2536	6	AR078193	Sequence
3	1328	92.9	1552	9	AF336436	Homo sapi
4	743.8	52.1	76504	9	AL355999	Human DNA
5	658.6	46.1	960	4	SSC271330	Sus scrofa
6	606	42.4	1310	9	BC058003	Homo sapi
7	588.2	41.2	1439	10	BC022734	Mus muscu
8	433	30.3	444	6	AX119033	Sequence
9	363.6	25.4	164125	9	AC018693	Homo sapi
10	362	25.3	143506	9	AC092833	Homo sapi
11	171.2	12.0	246357	2	AC109089	Rattus no
12	170	11.9	235407	2	AC094280	Rattus no
13	157.4	11.0	198115	2	AC099712	Mus muscu
14	147.6	10.3	165444	2	AC116285	Rattus no
15	105.6	7.4	321950	1	BX294136	Pirellula
16	102	7.1	13266	1	AE013749	Yersinia
17	102	7.1	205050	1	AJ414155	Yersinia
18	90	6.3	71464	9	AL513344	Human DNA
19	90	6.3	158917	2	AC016364	Homo sapi
20	86.8	6.1	9992	1	AE004255	Homo sapi
21	77.8	5.4	247450	1	AP005349	Vibrio ch
22	76.2	5.3	158917	2	AC016364	Vibrio ch
23	74.6	5.2	302331	1	AE016810	Mus muscu
24	70.8	5.0	198115	2	AC099712	Mus muscu
25	68.4	4.8	903	6	AR377690	Streptoco
26	55.2	3.9	4987	1	AE021365	Streptoco
27	53.4	3.7	10643	1	AE006492	Streptoco
28	53.4	3.7	13338	1	AE009972	Streptoco
29	53.4	3.7	50029	1	AE014140	Streptoco
30	53.4	3.7	311600	1	AP005141	Streptoco
31	52.4	3.7	1141	6	AX083744	Sequence
32	51.4	3.6	927	6	AR375248	Sequence
33	51.4	3.6	7218	6	I66494	Sequence
34	50.2	3.5	80471	2	AC139529	Botryllus
35	49.6	3.5	110000	2	PFMAL13_3	Continuation of 14
36	48.6	3.4	10049	1	AF291051	Carsonell
37	48.6	3.4	273275	3	AE014828	Plasmodi
38	48.4	3.4	8056	6	AX599046	Sequence
39	47.8	3.3	161773	2	AC119935	Mus muscu
40	47.4	3.3	212664	2	AC139241	Mus muscu
41	46.8	3.3	2000	6	AX655393	Sequence
42	46.4	3.2	153751	3	AC116551	Dictyoste
43	46.2	3.2	347582	3	PFMAL4P1	Plasmodi
44	45.6	3.2	1141	6	AX083744	Sequence
45	45.6	3.2	40068	3	CBRG44F10	Caenorhab

ALIGNMENTS

RESULT 1	BD270236	1429 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	Human glycosylation enzymes.				
DEFINITION	BD270236				
ACCESSION	BD270236.1	GI:33080004			
VERSION	JP 2002537796-A/3.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1429)				
AUTHORS	Coleman, R.A. and Betenbaugh, M.J.				
TITLE	Human glycosylation enzymes				
JOURNAL	Patent: JP 2002537796-A 3 12-NOV-2002;				

ORIGIN		/mol_type="unassigned DNA"	
Query Match		93.2%; Score 1332.4; DB 6; Length 2536;	
Best Local Similarity		94.3%; Pred. No. 0;	
Matches 1428; Conservative		0; Mismatches 1; Indels 85; Gaps 1;	
QY	1	ATGCGCTTCCCAAGAGAACTTTCAGGCTTTTGGCTGCAACCATCACCCCATGACT	60
DB	89	ATGCGCTTCCCAAGAGAACTTTCAGGCTTTTGGCTGCAACCATCACCCCATGACT	148
QY	61	GAGATGAGAGAAATCAACTTTTCAGTATTCGTAGTACGTGGATTATCTTGTGAAGAA	120
DB	149	GAGATGAGAGAAATCAACTTTTCAGTATTCGTAGTACGTGGATTATCTTGTGAAGAA	208
QY	121	CAGGAGTGAAGAACTTTTGTGAATGGCACACAGAGAAAGCCCTGTCCTGAGCGTC	180
DB	209	CAGGAGTGAAGAACTTTTGTGAATGGCACACAGAGAAAGCCCTGTCCTGAGCGTC	268
QY	181	TCAGAGCGTGCAGGTTGCAAGAGTGGGTGACAAAGGAGGAGCAAGCTGGATCAG	240
DB	269	TCAGAGCGTGCAGGTTGCAAGAGTGGGTGACAAAGGAGGAGCAAGCTGGATCAG	328
QY	241	GTGATAATTCAGTAGAGCACTGAGCTTGAAGAGTCAAGAGAACTGCGCCCAATGCA	300
DB	329	GTGATAATTCAGTAGAGCACTGAGCTTGAAGAGTCAAGAGAACTGCGCCCAATGCA	388
QY	301	GCAGAAATAGAGCTGATGGCATCGCTCATTTGCACCGTTCTTCTCAAGCCATGGACC	360
DB	389	GCAGAAATAGAGCTGATGGCATCGCTCATTTGCACCGTTCTTCTCAAGCCATGGACC	448
QY	361	AAAGATATCTCGATTAATTTCTAAGGAAGTGGCTGTCGCGCCCTGCGCCATTT	420
DB	449	AAAGATATCTCGATTAATTTCTAAGGAAGTGGCTGTCGCGCCCTGCGCCATTT	508
QY	421	TATTACTATCATCTTCTGCTGATGAGGTAAGATTCGCTGAGGAGTTTGTGGAT	480
DB	509	TATTACTATCATCTTCTGCTGATGAGGTAAGATTCGCTGAGGAGTTTGTGGAT	568
QY	481	GGGATTCGATAGATCCCAACCTTCCAGGCTGAAATTCAGTGTACAGATCTCTCTTA	540
DB	569	GGGATTCGATAGATCCCAACCTTCCAGGCTGAAATTCAGTGTACAGATCTCTCTTA	628
QY	541	GACTTCGGGCAATGTTGATCAGATCCGAGCAACATTTGCTTCTTTTGGGGTG	600
DB	629	GACTTCGGGCAATGTTGATCAGATCCGAGCAACATTTGCTTCTTTTGGGGTG	688
QY	601	GATGAGCAACTGTTGAGTGTCTGCTGATGGAGCAACTGGAGCAGTGGG	650
DB	689	GATGAGCAACTGTTGAGTGTCTGCTGATGGAGCAACTGGAGCAGTGGG	748
QY	651	-----	650
DB	749	AACTACTGGGAAAAGAACAAACAGATGTTGGAGGCTTTTGAACAAAGGACTTCTCT	808
QY	651	-----	695
DB	809	TTAGCCCTGAACTATCAGTTTGTATCCAGAGATTATCAACTTTGTTCTCAAACTAGGT	868
QY	696	TTTGAGTGTACAGCAACCAATCATGACTCTGCTCTGGGATTCGAATGGGCCA	755
DB	869	TTTGAGTGTACAGCAACCAATCATGACTCTGCTCTGGGATTCGAATGGGCCA	928
QY	756	CCCCGGCTTCCACTGAGAAACCTCCAGGAGTTTACTGATAGTGTGAAGCTAAACTG	815
DB	929	CCCCGGCTTCCACTGAGAAACCTCCAGGAGTTTACTGATAGTGTGAAGCTAAACTG	988
QY	816	AAGAGCTGGATTTCTTCTTCTTCTGATTTAAAGGATGGAACTTGGAGCTGTAGC	875
DB	989	AAGAGCTGGATTTCTTCTTCTTCTGATTTAAAGGATGGAACTTGGAGCTGTAGC	1048
QY	876	TAGTGCCTCTCTATCAAAATCAGGGTTTGACCTTGGAGACATAATCTACCTTAAATAGTGC	935
FEATURES		source	
1..1552		/organism="Homo sapiens"	
USA		Location/Qualifiers	
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			
Sood, R., Bonner, T. I., Makalowska, I., Stephan, D. A., Robbins, C. M., Connors, T. D., Morgenbesser, S. D., Su, K., Faruque, M. U., Pinkett, H., Graham, C., Baxeianis, A. D., Klinger, K. W., Landes, G. M., Carpten, J. D. and Trent, J. D.			
Cloning and characterization of 13 novel transcripts and the human RG88 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
Direct Submission			
Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892,			
USA			
1..1552			
/organism="Homo sapiens"			
Genomics 73 (2), 211-222 (2001)			
21218927			
11318611			
REFERENCE 2 (bases 1 to 1552)			

```

/mol_type="mRNA"
/db_xref="taxon:9606"
45..1007
/note="cl12 protein"
/codon_start=1
/product="C1orf13"
/protein_id="AAK25795.1"
/db_xref="GI:13430285"
translation="MAFPPKQLQGLVAATITPMIENGINSFVIGQYVDYLVEQGVK
NIFWNTGGTGLSLSVSERQVAEEWITGDKDQVITHVGALSKSQELAHQAEE
IGAGIAVIAFPFKIPKPTKDI--NFKLVAAAPALPFYVYHPALPTGVKLRAEELL
GILDKITFQSLGSDTDLDDGGCVDNQQQPAFLFGVDQLLSALVWGMATGAVGS
VYVHGKTNQMLBAFQESQDSALNVQFCIFQTFNFVVLGAFGVSQTKAIMTLVSGI
DMGDDRLIKQASQSEDSATAKISKIADFISPTDIDKQGNLEAGS"

```

ORIGIN

Query Match:	92.9%	Score 1328;	DB 9;	Length 1552;
Best Local Similarity	94.4%	Pred. No. 0;		
Matches 1423; Conservative	0;	Mismatches	3;	Indels 85; Gaps 1;
QY	1	ATGGCCCTCCCAAGAAAGAAACATTCAGGGCTCTGTGGCTGCAACCATCATCGCCCAATGACT	60	
DB	45	ATGGCCCTCCCAAGAAAGAAACATTCAGGGCTCTGTGGCTGCAACCATCATCGCCCAATGACT	104	
QY	61	GAGAAATCGAGAAATTCACACTTTTCAGTAAATTCGTCAGTATGTGGATTAATCTGTGAAAGAA	120	
DB	105	GAGAAATCGAGAAATTCACACTTTTCAGTAAATTCGTCAGTATGTGGATTAATCTGTGAAAGAA	164	
QY	121	CAGGGAGTGAGAACATATTTTGTCAATGGCAACACAGAGAGAGCCCTGTCCCTGAGGCTC	180	
DB	165	CAGGGAGTGAGAACATATTTTGTCAATGGCAACACAGAGAGAGCCCTGTCCCTGAGGCTC	224	
QY	181	TCAGAGGGTCGCCAGGTTCGACAGGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAG	240	
DB	225	TCAGAGGGTCGCCAGGTTCGACAGGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAG	284	
QY	241	GTGATTAATTCAGGTAGGAGCACTCAGCTTGAAGGAGTCACAGGAACTGGGCCCAACATGCA	300	
DB	285	GTGATTAATTCAGGTAGGAGCACTCAGCTTGAAGGAGTCACAGGAACTGGGCCCAACATGCA	344	
QY	301	GCAGAAATAGGAGCTGATGGGCATCGCTCTCATTCGACCGTTCTTCTCTCAAGCCCATGGACC	360	
DB	345	GCAGAAATAGGAGCTGATGGGCATCGCTCTCATTCGACCGTTCTTCTCTCAAGCCCATGGACC	404	
QY	361	AAAGATATCTCGATTAATTTCTCTAAAGGAGTGGCTGCTGCGCCCGCTGCCCTGCCATTT	420	
DB	405	AAAGATATCTCGATTAATTTCTCTAAAGGAGTGGCTGCTGCGCCCGCTGCCCTGCCATTT	464	
QY	421	TATTACATCATCTCCCTTGACAGGGGTAAAGATTCGTGCTGAGGAGTTGTTGGAT	480	
DB	465	TATTACATCATCTCCCTTGACAGGGGTAAAGATTCGTGCTGAGGAGTTGTTGGAT	524	
QY	481	GGGATCTGGATAAGATCCCAACCTTCGAAGGCTGAAATTCAGTGATACAGATCTCTTA	540	
DB	525	GGGATCTGGATAAGATCCCAACCTTCGAAGGCTGAAATTCAGTGATACAGATCTCTTA	584	
QY	541	GACTTCGGGCATGTGTTGATCAGAAATCCCAAGGCTGAAATTCAGTGATACAGATCTCTTA	600	
DB	585	GACTTCGGGCATGTGTTGATCAGAAATCCCAAGGCTGAAATTCAGTGATACAGATCTCTTA	644	
QY	601	GATGAGCAACTGTTCAGTGTCTCTGGTGATGGGAGCAACTGGAGCAGTGGG	650	
DB	645	GATGAGCAACTGTTCAGTGTCTCTGGTGATGGGAGCAACTGGAGCAGTGGGAGTACCTAT	704	
QY	651	-----	650	
DB	705	AACCTACCTGGGAAAHAAAGACAAACACAGATGTTGGAGGCTTTTGGACAAAGGACCTTCTCT	764	
QY	651	-----	650	
DB	765	TTAGCCCTGAACTATCATGTTTGTATCCAGAGATTTATCAACTTTGTTGTCAAACCTAGGT	824	
QY	696	TTTGGAGTGTACAGAGCAAAAGGACATCAAGCATCTGGTCTCTGGATTCGATGGGCCCA	755	

825	TTTGGAGTGTCA	CAGACCAAGCCATCATGACTCTGGTCTCTGGGATTC	CAATGGGCCCA	884
756	CCCGCGTCCACT	CGAGAAAGCCCTCCAGGAGTTTACTGATAGTCTGGAAGCT	AAACTG	815
885	CCCGCGTCTCCACT	CGAGAAAGCCCTCCAGGAGTTTACTGATAGTCTGGAAGCT	AAACTG	944
816	AAGAGCCTCGAATTC	CTTTCTTTCTCACTGATTTAAAGGATGGAAACTT	TGGAAGCTGGTAGC	875
945	AAGAGCCTCGAATTC	CTTTCTTTCTCACTGATTTAAAGGATGGAAACTT	TGGAAGCTGGTAGC	1004
876	TAGTGCTCTCTAT	CAAAATCAGGTTTGACCTTGAGCATATCTACCTTTAAATAGTGC		935
1005	TAGTGCTCTCTAT	CAAAATCAGGTTTGACCTTGAGCATATCTACCTTTAAATAGTGC		1064
936	ATCTTTTCTTCAGGGAATTT	TAGATCAACTCGAATAAACTCTCCTAGCAAAATCA	AAATCTC	995
1065	ATTTTCTTCAGGGAATTT	TAGATCAACTCGAATAAACTCTCCTAGCAAAATCA	AAATCTC	1124
996	ACAATAAGCATTTAG	STACCTTTTGTGAGCCTTAAAGAGTCTTATTTTGTGGAAGGGCAA		1055
1125	ACAATAAGCATTTAG	STACCTTTTGTGAGCCTTAAAGAGTCTTATTTTGTGGAAGGGCAA		1184
1056	AAACTCTAGAGTCA	CAACTCTCAGTCTCATTTCCACAGATTTTTTTTGTGGAGAATTT		1115
1185	AAACTCTAGAGTCA	CAACTCTCAGTCTCATTTCCACAGATTTTTTTTGTGGAGAATTT		1244
1116	CTGTTTATATGGATGAAAT	GGAAATCAAGAGGAAAATTTGTAATTGATTAATTC	CAATCTGTC	1175
1245	CTGTTTATATGGATGAAAT	GGAAATCAAGAGGAAAATTTGTAATTGATTAATTC	CAATCTGTC	1304
1176	TTTAGGAGCTCTCAT	TATATCTGGTCTGGTCTGTAATTCCTATTTTAAAGTTGCTCTAAT		1235
1305	TTTAGGAGCTCTCAT	TATATCTGGTCTGGTCTGTAATTCCTATTTTAAAGTTGCTCTAAT		1364
1236	TTAAACACATATAAT	GTCTTCAATTTAATAATTTCAATTTGGAAATCTAGGAAAAC	ACTC	1295
1365	TTAAACACATATAAT	GTCTTCAATTTAATAATTTCAATTTGGAAATCTAGGAAAAC	ACTC	1424
1296	TGAGCTACTGCATTTAG	CGAGGCACTTTAATACCAAACCTGAACATGCTCAACTGTGATA		1355
1425	TGAGCTACTGCATTTAG	CGAGGCACTTTAATACCAAACCTGAACATGCTCAACTGTGATA		1484
1356	CAACTCAAAATACAC	CAAGCTCAATTTGCTCTCAGTCTAACTCTAGAAATCGATGCTTTTG		1415
1485	CAACTCAAAATACAC	CAAGCTCAATTTGCTCTCAGTCTAACTCTAGAAATCGATGCTTTTG		1544
1416	AATTCATT	1423		
1545	AATTCATT	1552		

RESULT 4

AL355999	AL355999	76504 bp	DNA	linear	PRI 23-AUG-2001			
LOCUS	Human DNA sequence from clone Rp11-24906 on chromosome 1, complete sequence.							
DEFINITION	Human DNA sequence from clone Rp11-24906 on chromosome 1, complete sequence.							
ACCESSION	AL355999							
VERSION	AL355999.9	GI:15384805						
KEYWORDS	HTG.							
SOURCE	Homo sapiens (human)							
ORGANISM	Homo sapiens							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
	1 (bases 1 to 76504)							
REFERENCE	Tracey, A.							
AUTHORS	Direct Submission							
TITLE	Submitted (23-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,							
JOURNAL	CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk							
	Clone requests: clonerequest@sanger.ac.uk							
	On Aug 31, 2001 this sequence version replaced gi:13990035.							
	During sequence assembly data is compared from overlapping clones.							
	Where differences are found these are annotated as variations							
COMMENT								

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> Rpl1-24906 is from the library RPL1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone RPL1-24906. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone RPL1-24906 is at 76504 in this sequence. The true left end of clone RPL1-333G16 is at 55978 in this sequence. The true right end of clone RPL1-71D4 is at 2000 in this sequence.

FEATURES		Location/Qualifiers	
source		76504	
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
/chromosome="1"			
/clone="RP11-24906"			
/clone_lib="RPC1-11.1"			
ORIGIN			
Query Match 52.1%; Score 743.8; DB 9; Length 76504;			
Best Local Similarity 98.4%; Pred. No. 5.2e-180;			
Matches 751; Conservative 0; Mismatches 12; Indels 0; Gaps 0;			
QY	557	GATTATCAACTTTGTGTCAAACTAGGTTTGGAGTGTACAGACCAAGCCATCATGA 726	
DB	20322	GTTTCTCATTTGTTTTTCCCGTCTAGGTTTGGAGTGTACAGACCAAGCCATCATGA 20381	
QY	727	CTCTGGTCTCTGGATTCGAATGGGCCCCCGGCTTCCCTGTCAGAAAGCCCTCCAGGG 786	
DB	20382	CTCTGGTCTCTGGATTCGAATGGGCCCCCGGCTTCCCTGTCAGAAAGCCCTCCAGGG 20441	
QY	787	AGTTTACTGATGTGCTGAAGCTAAACTGAAGAGCCCTGGATTTCTCTTTCTACTGATT 846	
DB	20442	AGTTTACTGATGTGCTGAAGCTAAACTGAAGAGCCCTGGATTTCTCTTTCTACTGATT 20501	
QY	847	TAAAGGATGAACCTTGGAGCTGTAGTGTAGTGCCTCTCTATCAAACTCAGGTTTGCAC 906	
DB	20502	TAAAGGATGAACCTTGGAGCTGTAGTGTAGTGCCTCTCTATCAAACTCAGGTTTGCAC 20561	
QY	907	CTTGAGACATACTCTACTTTAAACTAGTGCATTTTCTTCTCAGGAAATTTTAGATGAACCT 966	
DB	20562	CTTGAGACATACTCTACTTTAAACTAGTGCATTTTCTTCTCAGGAAATTTTAGATGAACCT 20621	
QY	967	GAATAAATCTCTCTAGCAATGAATCTCACAATAAGCAATGAGTACCTTTTGTGAGCC 1026	
DB	20622	GAATAAATCTCTCTAGCAATGAATCTCACAATAAGCAATGAGTACCTTTTGTGAGCC 20681	
QY	1027	TTAAAAGTCTTATTTGTGAAGGGGCAAAACTCTAGGAGTCAAACTCTCACTCACTTC 1086	
DB	20682	TTAAAAGTCTTATTTGTGAAGGGGCAAAACTCTAGGAGTCAAACTCTCACTCACTTC 20741	

QY	1087	ATTTCACAGATTTTTCGTGGAGAAATTCGTTTATATGGATGAATGAATCAAGAG 1146	
DB	20742	ATTTCACAGATTTTTCGTGGAGAAATTCGTTTATATGGATGAATGAATCAAGAG 20801	
QY	1147	AAAATGTAATGATTAATTCATCTGCTTTAGGAGCTCTCATTTATCTCGGCTCTGGT 1206	
DB	20802	AAAATGTAATGATTAATTCATCTGCTTTAGGAGCTCTCATTTATCTCGGCTCTGGT 20861	
QY	1207	TCCATCTCCTATTTTAAAGTTCCTAAATTTAAACCACTATATATGCTTCAATTTAAT 1266	
DB	20852	TCCATCTCCTATTTTAAAGTTCCTAAATTTAAACCACTATATATGCTTCAATTTAAT 20921	
QY	1267	AATATTTCAATTTGGAACTCTAGGAACTCTGAGCTACTGCAATTTAGCGAGCACTTTAAT 1326	
DB	20922	AATATTTCAATTTGGAACTCTAGGAACTCTGAGCTACTGCAATTTAGCGAGCACTTTAAT 20981	
QY	1327	ACCAAACTGTAACTGTCTCACTGTATACAACTCAAAATACCAAGCTCATTGGCTGC 1386	
DB	20982	ACCAAACTGTAACTGTCTCACTGTATACAACTCAAAATACCAAGCTCATTGGCTGC 21041	
QY	1387	TCAGTCTAACTCTAGAACTGATGCTTTTGAATTCATTTCTGATG 1429	
DB	21042	TCAGTCTAACTCTAGAACTGATGCTTTTGAATTCATTTCTGATG 21084	

RESULT 5	SSC271330	960 bp	mRNA	linear	MAM 01-FEB-2001
LOCUS	Sus scrofa mRNA for acylneuraminase lyase (npl gene).				
DEFINITION	Sus scrofa mRNA for acylneuraminase lyase (npl gene).				
ACCESSION	AJ271330				
VERSION	AJ271330.1 GI:12657443				
KEYWORDS	acylneuraminase lyase; npl gene.				
SOURCE	Sus scrofa (pig)				
ORGANISM	Sus scrofa				
REFERENCE	1 Bruse, P., Traving, C. and Schauer, R.				
AUTHORS	Unpublished				
JOURNAL	2 (bases 1 to 960)				
REFERENCE	Bruse, P.				
AUTHORS	Direct Submission				
TITLE	Submitted (14-JAN-2000) Bruse P., Biochemisches Institut,				
JOURNAL	Christian-Albrechts-Universitaet, Olshausenstrasse 40, Kiel 24098,				
COMMENT	Germany				
	Alignment with the acylneuraminase lyase cDNA sequence from pig				
	kidney				
	(nplcna) enables the assembly of the following mouse partial ESTs:				
	gbAI550577, gbAA162738, gbW14179, gbAA544828, dbJAV158604,				
	dbJAV170531,				
	gbAA547408, and of the following human partial ESTs: gbW79930,				
	gbR97395,				
	gbAA226858, gbH65991, gbAI916625, gbAI521193, gbAI079177, to				
	complete				
	open reading frames homologous to microbial acylneuraminase lyases.				

FEATURES		Location/Qualifiers	
source			
1..960			
/organism="Sus scrofa"			
/mol_type="mRNA"			
/db_xref="taxon:9823"			
/tissue_type="kidney"			
1..960			
gene		/gene="npl"	
1..960			
CDS		/gene="npl"	
/codon_start=1			
/product="acylneuraminase lyase"			
/protein_id="CAC27797.1"			
/db_xref="GI:12857444"			
/db_xref="GOA:Q9BEC7"			
/db_xref="SPTREMBL:Q9BEC7"			
/translation="MASPKKKLQGLVAATITPMTEHGEINFSVIGQVYDLVEQGVK			
NIFVNGTGGSLISISERQVAAEWVTKGNKLDQIVIHVHALSLSESQELQHAAK			

IGADGIAVIAAPFELKPMKNDNLINFLKEVAAAAPALPFFYYHIFALTVKIVRAEELLD
GIQKPTTFGLKSESTDLDFGQCVDPQNDHQRFAPFLGVDQELLSALVNGATGAVGS
TYNLRKTNQMLEAFERKDFSSALNHQFCIQRFINFWKLGFGVSTKAIMTLVSGI
PMGPPLPLQKASREFDNAKAKLSLDVLSFTDLKXGNLEAC

ORIGIN		Query Match	Score	658.6;	DB 4;	Length	960;
		Best Local Similarity	83.4%;	Pred. No.	2.9e-158;		
		Matches	796;	Conservative	0;	Mismatches	74;
				Indels	85;	Gaps	1;
QY	1	ATGSCCTTCCCAAGAAACTTTCAGGCTCTGCTGCTGCAACCATCACCCCAATGACT	60				
DB	1	ATGSCCTTCCCAAGAAACTTTCAGGCTCTGCTGCTGCAACCATCACCCCAATGACT	60				
QY	61	GAGAAATCGAGAAATCAACTTTCAGTAAATGGTCAGTATGGAATATCTTGTGAAGAA	120				
DB	61	GAGCATCGAGAGATCAACTTTCAGTAAATGGTCAGTATGGAATATCTTGTGAAGAA	120				
QY	121	CAGGAGCTGAGAACATTTTGTGAATGCGCAACAGAGAGAGCCCTGCTCCCTGAGGCTC	180				
DB	121	CAGGAGCTGAGAACATTTTGTGAATGCGCAACAGAGAGAGCCCTGCTCCCTGAGGCTC	180				
QY	181	TCAGAGGCTCCAGGCTGAGAGAGAGTGGTGCACAAAGGAAGGACGACAGCTGATCAG	240				
DB	181	TCAGAGGCTCCAGGCTGAGAGAGAGTGGTGCACAAAGGAAGGACGACAGTTCATCA	240				
QY	241	GTGATAATTCAGTAGGAGCTAGCTTGAAGAGTCAAGAGACTGAGCACTGCCCAATGCA	300				
DB	241	ATAGTAATTCATGTAGGAGCTAGCTTGAAGAGTCAAGAGACTGAGCACTGCCCAATGCA	300				
QY	301	GCAGAAATGAGGCTGAGGAGCTGCTGATGACACCGTCTTCTTCAAGCCATGGACC	360				
DB	301	GCAGAAATGAGGCTGAGGAGCTGCTGATGACACCGTCTTCTTCAAGCCATGGACC	360				
QY	361	AAAGATATCTGATTAATTTCTAAAGGAAGTGGCTGTGCGGCCCTGCTCCCTGCCATTT	420				
DB	361	AAAGATATCTGATTAATTTCTAAAGGAAGTGGCTGTGCGGCCCTGCTCCCTGCCATTT	420				
QY	421	TATTACTATCATTCCTGCTTGAAGAGGCTTGAAGTTCGCTGAGGAGTGTGGAT	480				
DB	421	TATTACTATCATTCCTGCTTGAAGAGGCTTGAAGTTCGCTGAGGAGTGTGGAT	480				
QY	481	GGATATCTGATTAAGTCCACCTTCCAGGCTGAAATTCAGTGATACAGATCTCTTA	540				
DB	481	GGATATCTGATTAAGTCCACCTTCCAGGCTGAAATTCAGTGATACAGATCTCTTA	540				
QY	541	GACTTCGGGCAATGTGTGATCAGAAATCCAGCAACAGTTTGTCTTCTTTGGGGTG	600				
DB	541	GACTTCGGGCAATGTGTGATCAGAAATCCAGCAACAGTTTGTCTTCTTTGGGGTG	600				
QY	601	GATGAGCACTGTGAGTCTGCTGATGAGGAGCACTGGAGCAGTGGG-----	650				
DB	601	GATGAGCACTGTGAGTCTGCTGATGAGGAGCACTGGAGCAGTGGG-----	650				
QY	651	-----	650				
DB	651	-----	650				
QY	661	AACATCTGGGAGAAAGAACACAGATGTTGGAGGCTTTTGAACGAAGGACCTTCTCT	720				
DB	661	AACATCTGGGAGAAAGAACACAGATGTTGGAGGCTTTTGAACGAAGGACCTTCTCT	720				
QY	651	-----CAGTTTGTATCCAGAGATTTATCACTTGTGTCTCAACTAGGT	595				
DB	721	TCAGGCTTGAATCATCAGTTTTCATTAAGAGATTTATCACTTGTGTCTCAACTAGGT	780				
QY	696	TTTGGAGCTGTCAGACCAAGGACATCATCTGCTCTCTGGATTCCAAATGGGCCCA	755				
DB	781	TTTGGAGCTGTCAGACCAAGGACATCATCTGCTCTCTGGATTCCAAATGGGCCCA	840				
QY	756	CCCCGCTTCCACTGACAGAAAGCTCCAGGAGTTTACTGATAGTCTGAAAGCTAAACTG	815				
DB	841	CCCCGCTTCCACTGACAGAAAGCTCCAGGAGTTTACTGATAGTCTGAAAGCTAAACTG	900				
QY	816	AAAGAGCTGATTTCTTTCTTTCACTGATTTAAAGGATGGAACCTTGAAGCTGT	872				
DB	901	AAAGAGCTGATTTCTTTCTTTCACTGATTTAAAGGATGGAACCTTGAAGCTGT	957				

BC058003	1310 bp	mrna	linear	PRI 07-OCT-2003
LOCUS	Homo sapiens cDNA clone MGC:61869 IMAGE:4607228, complete cds.			
DEFINITION	Homo sapiens cDNA clone MGC:61869 IMAGE:4607228, complete cds.			
ACCESSION	BC058003			
VERSION	BC058003.1			
KEYWORDS	MGC.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	1 (bases 1 to 1310)			
AUTHORS	Srausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Rulyk, S.W., Villalón, D.K., Muzny, D.X., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Rodriguez, S., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, J., Smal, D.E., Scherf, A., Schein, J.E., Jones, S.J., and Marra, M.A.			
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
MEDLINE	22388257			
PUBMED	12477932			
REFERENCE	2 (bases 1 to 1310)			
AUTHORS	Srausberg, R.			
TITLE	Direct Substitution			
JOURNAL	Submitted 108-SEP-2003 National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT	Contact: MGC help desk Email: cgaphs-remail.nih.gov Tissue Procurement: CLONTECH cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc-mgcnhgri.nih.gov Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, C.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantrapop, S., Thomas, P.J., Touchman, J.W., Tsurgou, C., Vogt, J.L., Walker, N.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.			
FEATURES	Location/Qualifiers			
source	1..1310			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 48 Row: d Column: 12
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13540532.

```

/db_xref="taxon:3606"
/clone="MGC:61569 IMAGE:4607228"
/tissue_type="kidney"
/clone_lib="NIH MGC_75"
/lab_host="DH10B"
/note="Vec:or: PDNR-LIB"
109..811
/codon_start=1
/product="Unknown (protein for MGC:61869)"
/protein_id="AAH58003.1"
/db_xref="GI:34783662"
/translation="MAEPKKLQGLVAATIPMTENGINFSVIGQYDYLKVGK
NIFQVATIGESLISYSERQVAEWTKGDKLQVLIHVCALSLKESQELAAHAE
IGAGIATIAFFFLKPNKDLINFLKVAAPALPYHYHIALTKKRAEELLD
KLGKIPFQGLKSDTDLDFGQCVDQNRQQFAFEGDFEFCIORFINFVVKLENS
KLYSKNGRTPLGTFNPFLL"
124..717
misc_feature
/note="DapA; Region: Dihydropicolinate
synthase/N-acetylneuraminate lyase [Amino acid transport
and metabolism / Cell envelope biogenesis, outer
membrane]"
/db_xref="CDD:COG0329"

ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 9.9e-145; Length 1310;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCCCTTCCCAAGAAGAACTTCAGGGCTCTGTGGCTGCAACCATCATCGCAATGACT 60
Db 109 ATGGCCCTTCCCAAGAAGAACTTCAGGGCTCTGTGGCTGCAACCATCATCGCAATGACT 168
QY 61 GAGATGAGAAATCACTTTTCAGTAATGTGTGATGTGTGATATCTTGTGAAGAA 120
Db 169 GAGATGAGAAATCACTTTTCAGTAATGTGTGATGTGTGATATCTTGTGAAGAA 228
QY 121 CAGGAGTGAAGAACTTTTGTGAATGACACACAGGAGAGGCGCTGCTTGCAGCGTC 180
Db 229 CAGGAGTGAAGAACTTTTGTGAATGACACACAGGAGAGGCGCTGCTTGCAGCGTC 288
QY 181 TCAGAGCGTCGCGCAGGTTCAGAGAGAGTGGGTGACAAAAGGAGAGAGCAAGCTGGATCAG 240
Db 289 TCAGAGCGTCGCGCAGGTTCAGAGAGAGTGGGTGACAAAAGGAGAGAGCAAGCTGGATCAG 348
QY 241 GTGATAATTCAGTAGGAGACATGAGCTTGAGGAGCTCAGAGAACTGGCCCAACATGCA 300
Db 349 GTGATAATTCAGTAGGAGACATGAGCTTGAGGAGCTCAGAGAACTGGCCCAACATGCA 408
QY 301 GAGAAATAGGAGCTGATGCGATCGCTGTCATTGCAACCGTCTTCTTCAAGCCATGGACC 360
Db 409 GAGAAATAGGAGCTGATGCGATCGCTGTCATTGCAACCGTCTTCTTCAAGCCATGGACC 468
QY 361 AAAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCGCCCGCTGCGCTGCGCATTT 420
Db 469 AAAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCGCCCGCTGCGCTGCGCATTT 528
QY 421 TATTACTATCATATCTGCTTGAAGAGTGAAGATTCGCTGAGAGATGTTGGAT 480
Db 529 TATTACTATCATATCTGCTTGAAGAGTGAAGATTCGCTGAGAGATGTTGGAT 588
QY 481 GGGATTCGATGAAGAATCCCAACCTTCCAGAGGCTGAATTCAGTATGATACAGATCTCTTA 540
Db 589 GGGATTCGATGAAGATCCCAACCTTCCAGAGGCTGAATTCAGTATGATACAGATCTCTTA 648
QY 541 GACTTCGGCAATGTGTGATCAGATTCGACCAACAGATGCTTTCCTTTTGGGGTG 600
Db 649 GACTTCGGCAATGTGTGATCAGATTCGACCAACAGATGCTTTCCTTTTGGGGTG 708
QY 601 GATGAG 606
Db 709 GATGAG 714

```

```

RESULT 7
EC022734
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT

```

```

EC022734 1439 bp mRNA linear ROD 06-OCT-2003
Mus musculus N-acetylneuraminate pyruvate lyase, mRNA (cDNA clone
MGC:31604 IMAGE:4506776), complete cds.
EC022734
EC022734.1 GI:18490966
MGC.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1439)
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
Hepkins R.F., Jordan H., Moore T.I., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L.,
Schetz T.E., Brownstein M.J., Ustin T.B., Tschiyuki S.,
Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J.,
Abramson R.D., Mullahy S.J., Sosak S.A., McEwan P.J.,
McKernan K.J., Malle J.A., Gunaratne P.H., Gay L.J., Hulyk S.W.,
Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalobos D.K., Muzny D.M., Sodergren E.C., Lu X., Gibbs R.A.,
Fahy J., Helton E., Kettner M., Madan A., Rodriguez S.,
Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y.,
Bouffard G.G., Blakesley R.W., Tsachman J.W., Green E.D.,
Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Bickertfield Y.S., Krzywinski M.I., Skalska U., Smalinski D.E.,
Scherch A., Schein J.E., Jones S.J. and Marra M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 15999-16003 (2002)
22388257
12477932
2 (bases 1 to 1439)
Strausberg R.
Direct Submission
Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-259C,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 44 Row: h Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21311854.
Location/Qualifiers
1..1439
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="MGC:31604 IMAGE:4506776"
/tissue_type="Eye, retina, mouse strain C57BL/6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..1439
/gene="Npl"

```

FEATURES
source

gene


```
QY 1 ATGGCTCTCCCAAGAGAACTTCAAGGCTTTGTGGCTGCAACCAATCAGGCCAATGACT 60
Db 1 ATGGCTCTCCCAAGAGAACTTCAAGGCTTTGTGGCTGCAACCAATCAGGCCAATGACT 60
QY 61 GAGATGGAGAAATCACTTTTCAGTAATGGTCACTATGTGATTAATCTTGTGAAGAA 120
Db 61 GAGATGGAGAAATCACTTTTCAGTAATGGTCACTATGTGATTAATCTTGTGAAGAA 120
QY 121 CAGGGAGTGAAGAACATTTTGTGAATGACACACAGGAGAGGCGCTGCTCCCTGAGCGTC 180
Db 121 CAGGGAGTGAAGAACATTTTGTGAATGACACACAGGAGAGGCGCTGCTCCCTGAGCGTC 180
QY 181 TCAGACGTCGCCAGCTGCAGAGAGTGGGTCAAAAAGGGHAGGACAGCTGGATCAG 240
Db 181 TCAGACGTCGCCAGCTGCAGAGAGTGGGTCAAAAAGGGHAGGACAGCTGGATCAG 240
QY 241 GTGATAATTCACGTAGAGACACTGAGCTTGAAGAGTCAAGGAACTGGCCCAACATGCA 300
Db 241 GTGATAATTCACGTAGAGACACTGAGCTTGAAGAGTCAAGGAACTGGCCCAACATGCA 300
QY 301 CGAGAAATAGGAGCTGACGGCATCGCTGATTCATTCGACCGTCTTCTCTCAAGCCATGGACC 360
Db 301 CGAGAAATAGGAGCTGATGGCATCGCTGATTCATTCGACCGTCTTCTCTCAAGCCATGGACC 360
QY 361 AAGATATCCTGATTAATTTCTTAAAGGAGTGCTCTCGCCGCCCTGCCCTGCCATTT 420
Db 361 AAGATATCCTGATTAATTTCTTAAAGGAGTGCTCTCGCCGCCCTGCCCTGCCATTT 420
QY 421 TATTACTATCACATTCCTGCTTGA 445
Db 420 TATTACTATCACATTCCTGCTTGA 444

RESULT 9
AC018693 164125 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-555N21 from 2, complete sequence.
DEFINITION AC018693
ACCESSION AC018693.8 GI:15321565
VERSION AC018693.8
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Suietion,J.B. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
8477074
REFERENCE 2 (bases 1 to 164125)
AUTHORS Belter,E., Cotton,M. and Doebber,A.
TITLE The sequence of Homo sapiens BAC clone RP11-555N21
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 164125)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 164125)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 164125)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 28, 2001 this sequence version replaced gi:14141805.
```

```
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0555N21
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phased quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Fietz de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP13-490C19; the clone sequenced to the right is RP11-173C1, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-555N21; actual end is at base position 31977 of RP11-173C1.

Polymorphisms exist between RP11-555N21 and RP11-173C1.

FEATURES	Location/Qualifiers
source	1. .164125 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="2" /map="2"
misc_feature	/clone="RP11-555N21" 1128. .1542 /notes="similar to Homo sapiens EST AV756546 (NID:910914394)"
misc_feature	1136. .1543 /notes="similar to Homo sapiens EST A1123298 (NID:93539064); ov48C08.x1" 1137. .1348 /notes="similar to Homo sapiens EST BF573448 (NID:911647160)"
misc_feature	1138. .1482 /notes="similar to Homo sapiens EST BF695725 (NID:911981133)"
misc_feature	1175. .1542 /notes="similar to Homo sapiens EST BG942016 (NID:914341388) ax20a10.x1"
misc_feature	1180. .1479 /notes="similar to Homo sapiens EST AV763398 (NID:910921246)"
misc_feature	1182. .1543

```
/note="similar to Homo sapiens EST AW156903 (NID:g5228304)
au89d05.x1"
misc_feature
1274..1556
/note="similar to Homo sapiens EST AA483370 (NID:g2212183)
ne74c02.s1"
misc_feature
1274..1553
/note="similar to Homo sapiens EST AA614576 (NID:g2466772)
np53d10.s1"
misc_feature
1274..1542
/note="similar to Homo sapiens EST AA722249 (NID:g2739956)
zh21f04.s1"
repeat_region
1533..1553
/rpt_family="AT-rich"
repeat_region
2087..2396
/rpt_family="Alu"
repeat_region
2367..2399
/rpt_family="(A)n"
misc_feature
2373..2398
/note="similar to Homo sapiens EST BG942015
(NID:g14341388) ax20a10.x1"
misc_feature
2379..2401
/note="similar to Homo sapiens EST AV756546
(NID:g10914394)"
repeat_region
2612..2792
/rpt_family="MIR"
repeat_region
2799..2837
/rpt_family="(CTA)n"
repeat_region
3039..3105
/rpt_family="L2"
repeat_region
3168..3467
/rpt_family="Alu"
repeat_region
3604..3829
/rpt_family="MaLR"
repeat_region
3990..4294
/rpt_family="Alu"
repeat_region
4313..4403
/rpt_family="ERV1"
repeat_region
4404..4706
/rpt_family="Alu"
repeat_region
4707..4769
/rpt_family="ERV1"
misc_feature
4784..5032
/note="similar to Homo sapiens EST BF876452
(NID:g12266582)"
misc_feature
4811..5032
/note="similar to Homo sapiens EST AL525867
(NID:g12789360)"
misc_feature
4926..5032
/note="similar to Bos taurus EST BF653570 (NID:gl1318702)"
repeat_region
5038..5069
/rpt_family="(T)n"
repeat_region
5040..5351
/rpt_family="Alu"
misc_feature
5058..5076
/note="similar to Homo sapiens EST AV728536
(NID:g10837957)"
misc_feature
5164..5166
/note="similar to Homo sapiens EST BF740024
(NID:g12066700)"
misc_feature
5353..5725
/note="similar to Homo sapiens EST AL554919
(NID:g12896160)"
misc_feature
5358..5902
/note="similar to Homo sapiens EST AL577521
(NID:g12940733)"
misc_feature
5362..5687
/note="similar to Homo sapiens EST AL525867
(NID:g12789360)"
misc_feature
5362..5498
/note="similar to Homo sapiens EST BF876452
(NID:g12266582)"
misc_feature
5364..5856
/note="similar to Homo sapiens EST BF740024
```

```
(NID:g-2066700)"
5364..5523
/note="similar to Bos taurus EST BF653570 (NID:g-1918702)*
5457..5687
/note="similar to Homo sapiens EST H65991 (NID:gl024731)
yr73b09.r1"
misc_feature
5473..5725
/note="similar to Homo sapiens EST AW192585 (NID:g6471284)
xl47a06.x1"
misc_feature
5548..5725
/note="similar to Homo sapiens EST AV728536
(NID:g10837957)"
misc_feature
5569..5725
/note="similar to Homo sapiens EST AI245446 (NID:g3840843)
qk29f09.x1"
misc_feature
5637..5725
/note="similar to Homo sapiens EST AI916625 (NID:g5656480)

Query Match 25.4%; Score 363.6; DB 9; Length 164.25;
Best Local Similarity 83.0%; Pred. No. 3.3e-82;
Matches 478; Conservative 0; Mismatches 84; Indels 14; Gaps 5;

QY 487 CTGGATAGATCCACACCTTCCAGGGCTGAATTCAGTACAGATCTCTAGACTTC 546
Db 5347 CCGGCCAAGAGCCCACTTTCTAAAGGCTGAATTCAGTACAGACA--TCCTAGACTTC 5404
QY 547 GGGCAATGTTGATCAGAAATGCCAGCAACAGTTTGCTTCTTTTGGGGTGGATGAG 606
Db 5405 AGGCAATGTTGATCAGAAATGCCAGCAATAGTTGACCTTCTTTGGAGGGTGGATGAG 5464
QY 607 CAACTGTTGAGTGTCTGTGTGATGGAGCAACTGGAGAGTGGGAGTTTGTATCAGA 666
Db 5465 CAACTGTTGAGTGTCTGTGTGATGGAGCAACTGGAGAGTGGGAGATTATATCGACA 5524
QY 667 GATTTATCAACTTTGTTGTCAACTAGTTTGGAGTGTACAGACCAAGCCATCATGA 726
Db 5525 GATTTATAAATTTTGTAGTCAAACTAGTTTCTAGTGTACAGATCAAAAGCCATCATGA 5584
QY 727 CTCTGTCCTCTGGGATCCAAAGGGCCCAACCCGGCTTCCACTGCAGAAAGCCCTCAGGG 786
Db 5585 CTCTGTCCTCTGGAGTGGAGTGGGCCCCACCCATCTTCGGTGCAGAAAGCCCTCAGGG 5644
QY 787 AGTTTACTGATAGTGTGAAGCTAAACTGAAGAGCCTGGAATTCCTTTCTTCTCACTGATT 846
Db 5645 AGTTTACTGATAGTGTGAAGCCAACTAAAGAGCTTGGATT--TTTCTTCTCACTGGCT 5703
QY 847 TAAAGATGGAATCTGGAGCTGGTAGTGTAGTGTCTCTATCAATCAGGGTTTGCAC 906
Db 5704 TAAAGATGGAATCTGGAGCCTATAGCCAGTGTCTATCCCAATCAGGCTGTGTGC 5763
QY 907 CTTGACACATATCTACCTTAAATAGTGCATTTTTTTTCTCAGGGA--ATTTAGATGAAT 965
Db 5764 ATTAGACATAGTCTGCCCTTAATAGTGTGATCTTTTCTCAGGGACTTTTAGAAAACT 5823
QY 966 TGA---TAACTCTCTAGCAAAATGAAATCTCA-----CAATGACATTCAGGTACC 1015
Db 5824 TGAACCTGAATCTCTCTAGCAAAATGAAATCTCAATCTCAATCAACAGATATTTAAGTACT 5883
QY 1016 TTTTGTGACCTTAAAGTCTTATTTCTGAGGG 1051
Db 5884 TACTATGAGCCTTAAAGGTTTATTTTAGGCGGG 5919

RESULT 10
AC092833 143506 bp DNA linear PRI 29-MAY-2002
LOCUS Homo sapiens BAC clone RP13-490C19 from 2, complete sequence.
DEFINITION AC092833
ACCESSION AC092833
VERSION AC092833.4 GI:20331024
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```


Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

/db_xref="taxon:10116"
/clone="CH230-20223"
1. .1529
/note="wgs end extension
clone end:Sp6"
4261. .4701
/note="clone boundary
clone_end:Sp6"
site:
end_sequence:BH333658"
5717. .6227
/note="clone boundary
clone_end:T7"
site:
end_sequence:BH333657"
58725. .70851
/note="wgs end extension
clone end:T7"
71895. .73424
/note="wgs end extension
clone_end:T7"
ORIGIN
Query Match 12.0%; Score 171.2; DB 2; Length 246357;
Best Local Similarity 55.4%; Pred. No. 1.1e-32;
Matches 378; Conservative 0; Mismatches 293; Indels 11; Gaps 2;
QY 626 TGATGGAGCACTGAGGAGTGGGAGTTTCTTCCAGAGATTTATCACTTTGTTGT 685
Db 120203 TTAGGTGTCAATTTGTTCTTTGAGACGTTTGGAGAAATGTCCTCTATTTTC 120144
QY 686 CAACTAGTTTGGAGTGTCAAGACCAAGCCATCATGCTGCTCTCTGGGATTC 745
Db 120143 CCATCTAGTTTGGAGTGTGGAGCCAAAGCCATCATGCTGCTCTGGGATTC 120084
QY 746 AATGGGCCCAACCCCGGCTTCCACTGCAGAAAGCCTCCAGGAGTTTACTGATGTCGA 805
Db 120083 AATGGGCCCAACCCCGGCTTCCCTCTGCAGAAAGCCACCCAGGAGTTTACTGATGTCGA 120024
QY 806 AGCTAACTGAGAGAGCTGGATTTCTCTTCTTCTCACTGATTTAAAGATGGAACCTTGA 865
Db 120023 GGCCAGCTGAAGAGCTGAATTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 119964
QY 866 AGCTGAGTCTAGTCTGCTCTCTCTATCAATCAAGGTTTGCACCTTGCAGACATAATCTACCT 925
Db 119963 AGCTGAGTCTAGTCTGCTCTCTCTATCAATCAAGGTTTGCACCTTGCAGACATAATCTACCT 119913
QY 926 TAAATAGTCTCAATTTTCTCTAGGGAATTTTAGATGCACTTGAATAAATCTCTCTAGCAA 985
Db 119912 TTCTTTTCTCNGGGATTTTCTAGTGAATGTAAGTCTTCTTCTTCTTCTTCTTCTTCTTCT 119853
QY 986 ATGAATCTCAATGAAGATTTGAGTCTCTTCTGAGCTTTAAAGCTTTATTTTGT 1045
Db 119852 CCCACACTAATCAAGAAATTTTAAAGTATCAATATGAACCTTCTTCTTCTTCTTCTTCT 119793
QY 1046 GAAGGGGCAAACTCTAGGAGTCAACCTCTCAGTCAATCTTCAAGATTTTCTTCTTCTTCTTCT 1105
Db 119792 GGAGTCAAACTCTTCAAGAAATGAGGATCTTATGTAATGTAATGCTTCTTCTTCTTCTTCT 119733
QY 1106 GGAAGAAATTTCTGTTTATATGATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1165
Db 119732 TTTTAAATGCTTTTCCAAAGGAATCCATAACTGAAGTCAAGATATATGAAGCTTTAGT 119673
QY 1166 TCATCTCTCTTTAGGAGCTCTCTATCTCTGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 1223
Db 119672 TATCCAAAGAGAGAGAGCTTTAGGAATTTGCTGCTTCTGCTCTCTCTCTCTCTCTCTCTCT 119613
QY 1224 AGTTGCTAATTTTAAACCACTATATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1283
Db 119612 AACTTTTCAATAGGAGAAATAGATTTATCATCATAGATAGTAAAGATCTCTCTCTCTCTCT 119553
QY 1284 CTAGGAAACTCTGAGCTACTG 1305

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: Q2BP
Center clone name: CH230-202E3
Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 234872 bases at least Q40
Consensus quality: 236549 bases at least Q30
Consensus quality: 237840 bases at least Q20
Estimated insert size: 240127; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 65325: contig of 65325 bp in length
* 65326 65425: gap of unknown length
* 68624: contig of 3199 bp in length
* 68625 68724: gap of unknown length
* 68725 244850: contig of 176126 bp in length
* 244851 244950: gap of unknown length
* 244951 246357: contig of 1407 bp in length.
Location/Qualifiers
1. .246357
/organism="Rattus norvegicus"
/mol_type="genomic DNA"

FEATURES
source

Db 119552 CCTGAACACGAGTGTGTGA-G 119531

RESULT 12
AC094280
LOCUS AC094280 235407 bp DNA linear HTG 14-NOV-2002
DEFINITION Rattus norvegicus clone CH230-3G22, WORKING DRAFT SEQUENCE, 3
unordered pieces.
ACCESSION AC094280
VERSION AC094280.10 GI:24954121
KEYWORDS HTG; HTGS; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 235407)
Murny D, Marie, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, X., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
DeGado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Escoto-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Freaser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Guraratne, P., Haaland, M., Hamill, C., Hamilton, N., Hernandez, J.,
Harvey, Y., Haykal, P., Hawes, A., Henderson, N., Hamilton, K.,
Herrandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, D., Kowar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Lulseged, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Margum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Mirza, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwakoemeloh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, T., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, C.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, F., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, B., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhou, J., Zhao, S., Zeng, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G., and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 235407)
Worley, K.C.
Direct Submission
TITLE Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE
AUTHORS
JOURNAL

Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 235407)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (14-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 14, 2002 this sequence version replaced gi:2326f009.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAZH
Center clone name: CH230-3G22
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 226392 bases at least Q40
Consensus quality: 228011 bases at least Q30
Consensus quality: 228972 bases at least Q20
Estimated insert size: 230729; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved
* 1 232242: contig of 232242 bp in length
* 232243 232342: gap of unknown length
* 232343 233631: contig of 1289 bp in length
* 233632 233731: gap of unknown length
* 233732 235407: contig of 1676 bp in length.
----- Location/Qualifiers
1. 235407
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-3G22"
1. 1252
/note="wgs_end_extension"
clone_end:Sp6"
4523. 5686
/note="wgs_end_extension"
clone_end:Sp6"
5995. 7857
/note="wgs_end_extension"
clone_end:Sp6"
8387. 133612
/note="clone boundary"
clone_end:Sp6
site:
end_sequence:BH31006f"
complement(229906. 230730)
/note="clone_boundary"

FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```
ORIGIN
Query Match 11.9% Score 170; DB 2; Length 235407;
Best Local Similarity 78.7%; Pred. No. 2,1e-32;
Matches 203; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

clone_end:T7
site:
end_sequence: BH310064"

QY 626 TGATGGGACAACTGAGCAGTGGGAGTGTGATATCCAGAGATTATCAACTTTGTTT 685
DB 523 TTAGGTGTACAAATTTTCCTTTGAGAGCTTTGGGAGAAATGTCCTCACTCCCTATATTTTC 982
QY 686 CAAACTAGTTTTGGAGTGTCAAGACCAAGCCATCATCACTCTGTTCTGGGATGCC 745
DB 983 CCATCTAGTTTTGGAGTGTCCGAGACCAAGCCATCATCACTCTGTTCTGGGATGCC 1042
QY 746 AATGGGCCACCCCGCTCCACTGCAGAAAGCTCCAGGAGTTTACTGTAGTGTCTGA 805
DB 1043 AATGGGCCACCCCGCTTCCTCTGCAGAAAGCCACCCAGGAGTTTACTGTCTATATCTGA 1102
QY 806 ACCTAAACTGAGAGCTCGATTCCTTTCTTCTCACTGATTTAAAGGATGGAACCTTGA 865
DB 1103 GCGCAAGCTGAAGAGCTGAATTCCTTTCTTCTCCCTGTTTAAAGGATGGAACATGGA 1162
QY 866 AGCTGTAGCTAGTGCTT 883
DB 1163 AGCTGTAGCTAGTGCTT 1180

RESULT 13
AC099712 198115 bp DNA linear HTG 18-NOV-2001
LOCUS Mus musculus clone RP23-400C10, WORKING DRAFT SEQUENCE, 17
DEFINITION Unordered pieces.
ACCESSION AC099712
VERSION AC099712.1 GI:16974211
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 198115)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,B.
TITLE Mus musculus, clone RP23-400C10
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 198115)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,B., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeAtellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulse,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlegza,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.
Direct Submission
TITLE Submitted (18-NOV-2001) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1402: contig of 1402 bp in length
* 1403: gap of 100 bp
* 1503: contig of 1018 bp in length
* 2521: gap of 100 bp
* 4431: contig of 1811 bp in length
* 4531: gap of 100 bp
* 4532: contig of 1529 bp in length
* 6061: gap of 100 bp
* 6161: contig of 2651 bp in length
* 8821: gap of 100 bp
* 8822: contig of 2289 bp in length
* 11211: gap of 100 bp
* 11311: contig of 6510 bp in length
* 17820: gap of 100 bp
* 17920: gap of 100 bp
* 23045: contig of 5125 bp in length
* 23446: gap of 100 bp
* 23446: contig of 33641 bp in length
* 56786: gap of 100 bp
* 56887: contig of 10698 bp in length
* 67585: gap of 100 bp
* 67685: contig of 14035 bp in length
* 81720: gap of 100 bp
* 81820: contig of 20158 bp in length
* 101978: gap of 100 bp
* 102078: contig of 17037 bp in length
* 119115: gap of 100 bp
* 119215: contig of 20006 bp in length
* 139220: gap of 100 bp
* 139221: contig of 20677 bp in length
* 139221: gap of 100 bp
* 159998: contig of 32350 bp in length
* 160098: gap of 100 bp
* 192448: contig of 5568 bp in length.
* 192548: Location/Qualifiers
* 198115: /organism="Mus musculus"
* 198115: /mol_type="genomic DNA"
* 198115: /db_xref="taxon:10090"
* 198115: /clone="RP23-400C10"
* 198115: /clone_lib="RPC1-23 Female Mouse BAC"
* 198115: /note="assembly_fragment"
* 1503: .2520

misc_feature
misc_feature
```

189173 ATTATTA-TAGAGAAATCTACTATTATAGAGATGCATGACCCATGCGAAATT 189226

AC116285	165444 bp	DNA	linear	HTG 19-NOV-2002			
LOCUS	Rattus norvegicus clone CH230-362N7, WORKING DRAFT SEQUENCE, 2						
DEFINITION	unordered pieces.						
AC116285							
AC116285.4	GI:250731175						
VERSION	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.						
KEYWORDS	Rattus norvegicus (Norway rat)						
SOURCE	Rattus norvegicus						
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.						
REFERENCE	1 (bases 1 to 165444)						
AUTHORS	Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Gavazos, I., Ceasar, H., Center, A., Cracko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, X., Foster, P., Fraser, C. M., Gabisti, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregoire-Goris, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hices, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulseg, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, N., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathewney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Narkervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoeleneh, O., Okwodu, G., Olampong, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, D., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savory, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, S., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.						
REFERENCE	Direct Submission						
AUTHORS	Unpublished						
TITLE	2 (bases 1 to 165444)						
DEFINITION	Worley, K. C.						
KEYWORDS	Direct Submission						
SOURCE	Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One						

Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 165444)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:22856515.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GUCR
 Center clone name: CH230-362N7
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 158039 bases at least Q40
 Consensus quality: 159792 bases at least Q30
 Consensus quality: 160671 bases at least Q20
 Estimated insert size: 161405; sum-of-contigs estimation
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Gennack_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 164095: contig of 164095 bp in length
 * 164096 164195: gap of unknown length
 * 164196 165444: contig of 1249 bp in length.

FEATURES
 source
 Location/Qualifiers
 1..165444
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-362N7"
 misc_feature 1..1192
 /note="wgs_contig"

ORIGIN
 Query Match 10.3%; Score 147.6; DE 2; Length 165444;
 Best Local Similarity 68.8%; Pred. No. 1.2e-26;
 Matches 253; Conservative 0; Mismatches 99; Indels 16; Gaps 3;

QY 632 GAGCACTGGAGCTGGGAGTTTGTATCCAGAGATTATCACTTTCTGTCAACT 691
 Db 4-607 GAACAGAAGGCTCAGTCACCAAGTCTACATCCAGGATTTATCACTATG 41548
 QY 692 AGCTTTGGAGTGCAGACCAAGCCATCATGACTCTGGTCTCTGGGATCCAAATGGG 751
 Db 41547 AGCTTTGGAGTGCAGACCAAGCTATCATGACCTAGTCTCTGGGATCCAAATGGG 41488
 QY 752 CCCACCCCGGCTTCCACTGCGAGAAAGCCTCCAGGGAGTTTACTGATAGTCTGAAGCTAA 811

Db 41487 CCCACCCCGGCTTCCACTGCGAGAGTCCACCGAGAGTTTCTCTGCTAATG----- 41438
 QY 812 ACTGAAGAGCTGGATTTCTTTCTTCACTGATTTAAAGAGTGAAGACTGGAGCTGG 871
 Db 41437 -TTGAAGAGCTGGAATTTCCCTTCTTCCCTAGATTAAAGATGGGACTTGGAGACTG 41379
 QY 872 TACTAGTGGCTCTCTATCAAAATCAGGTTTGACCTTTGAGACATATCTACCTTAATA 931
 Db 41378 TACTGGTGC-----CTGATAGCTAGTGTGTAGCCGAGACTTTATCTATCTTCAATA 41323
 QY 932 GTCCATTTTCTCAGGGA-ATTTAGATGAACCTTGAATAAATCTCTCTAGCAATGAA 990
 Db 41322 GTCCATTTCTCTCAGGACTTTTAGGTGAATTTGACCAAGCTTTCTTTAGCAGTA 41263
 QY 991 ATCTCACA 998
 Db 41262 GAATCCCA 41255

RESULT 15
 EX294138/c
 LOCUS BX294138 321950 bp DNA linear BCT 11-JUL-2003
 DEFINITION Pirellula sp. strain 1 complete genome; segment 6/24.
 ACCESSION BX294138 EX119912
 VERSION BX294138.1 GI:32443466
 KEYWORDS complete genome.
 SOURCE Pirellula sp. 1
 ORGANISM Pirellula sp. 1
 Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
 Planctomycetaceae; Pirellula.
 REFERENCE 1 (Bases 1 to 321950)
 Gloeckner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T.,
 Ludwig, W., Gade, D., Beck, A., Borzym, K., Heitmann, K., Rabus, R.,
 Schlöter, H., Amann, R. and Reinhardt, R.,
 Complete genome sequence of the marine planctomycete Pirellula sp.
 strain 1
 Proc. Natl. Acad. Sci. U.S.A. 100 (14), 8298-8303 (2003)
 MEDLINE 23735913
 PUBMED 12835416
 REFERENCE 2 (Bases 1 to 321950)
 Kube, M., Borzym, K., Heitmann, K., Klages, S., Marguardt, I.,
 Leitzack, S., Beck, A., Pawlik, R., Reinhardt, R., Gloeckner, F.O.,
 Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Rabus, R.,
 Schleutner, H. and Amann, R.
 Direct Submission
 Submitted (21-JAN-2003) Max Planck Institute for Molecular
 Genetics, proScience Imnestrasse 73, D-14195 Berlin, Germany Max
 Planck Institute for Marine Microbiology Celsiusstrasse 1, D-28359
 Bremen, Germany
 This project was carried out by
 *Max Planck Institute for Molecular Genetics, Berlin, Germany; *Max
 Planck Institute for Marine Microbiology, Bremen, Germany; in the
 framework of the REGX-project, http://www.regx.de -----
 Genome Center
 Center: Max Planck Institute for Molecular Genetics
 Center code: MPIMG
 ----- Summary Statistics
 Sequencing vector: pUC19; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 7142841 bases at least Q40
 Consensus quality: 7145138 bases at least Q30
 Consensus quality: 7145484 bases at least Q20
 Quality coverage: 8.03

 This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid sequence; assembly was additionally confirmed by long
 range PCR and cosmid end sequences.

See <http://www.micro-genomes.mpg.de/pirellula/> for more information including minimal tiling path from a set of 220 cosmids out of 908. See the misc feature tag below for the boundaries of the MTP cosmids. ----- Annotation

Center: Max Planck Institute for Marine Microbiology
Celsiusstrasse 1, D-28359 Bremen, Germany.

Center Code: MEIMM
Email: fcg@mpi-bremen.de
Phone: +49 (0)421 2028 938 Fax: +49 (0)421 2028 580

Three different programs (Glimmer, Critica and Orpheus) were used for ORF-prediction. A nonredundant list of ORFs was generated by suitable parsing of the results.

Automated annotation was done with the software package Pedant Pro (<http://www.biomax.de>). All ORF predictions and annotations were manually corrected by considering all results of the different tools applied. See <http://www.regx.de> for more information and access to supplementary information.

FEATURES

source

Location/Qualifiers
1..321950

/organism="Pirellula sp. 1"

/mol_type="genomic DNA"

/strain="1"

/db_xref="taxon:243090"

complement(362..1051)

/locus_tag="RB2934"

complement(362..1051)

/locus_tag="RB2934"

/function="lipid, fatty-acid and isoprenoid biosynthesis; fatty acid biosynthesis; lipid, fatty-acid and isoprenoid metabolism"

/note="PMID: 2144277 PMID: 10784042 PMID: 11481430 best DB hits: BLAST: swissprot:O67611; ACP_AQUAE ACYL CARRIER PROTEIN (ACP) ----; E=0.018 swissprot:P19372; ACP_RHIME ACYL CARRIER PROTEIN (ACP) ----; E=0.055

swissprot:O9G22; ACP_RHILE ACYL CARRIER PROTEIN (ACP)

----; E=0.064 COG: aq_1217a; COG0236 Acyl carrier

protein; E=0.002 PFAM: PF00950; Phosphopantetheine

attachment site; E=0.3"

/codon_start=1

/transl_table=11

/product="similar to acyl carrier protein"

/protein_id="CAD72909.1"

/db_xref="GI:32443467"

/translation="NGLDAVEITVMIVEDHFGISINDETERVLTGVDLVALIQSRIGA
AEIATCPITLSERLSSVRELTNDQTLRIGTRVDMNRTQRLQWADDDILGT
RAGPLRPAILKRLAFLATTFVIAFLGSVTIDVAILPLTALAACATLALHTIP
FRSIPDAAATGAIARRMAGISVATKQLHLRTDSEILHELPRIVAATLTGDSKTR
TTRFTDLGMG"

complement(1184..1408)

/locus_tag="RB2936"

complement(1184..1408)

/locus_tag="RB2936"

/function="cellular communication/signal transduction"

/note="PMID: 8134129 best DB hits: BLAST:

swissprot:PS3356; HT16 HYDAT TYROSINE-PROTEIN KINASE HTK16

----; E=0.007 gb:AAF37305.1; AC005931_3 (AC005931)

6-Phosphofructo-2-kinase; E=0.008 gb:AAF58766.1;

AE003828) CG16728 gene product [Drosophila; E=0.009 PFAM:

PF00023; Ank repeat; E=2.4e-08"

/codon_start=1

/transl_table=11

/product="similar to tyrosine-protein kinase HTK16"

/protein_id="CAD72910.1"

/db_xref="GI:32443468"

/translation="MQAQLHPRRTFGVSPSANMLTQLMDHHRHTPLHEACTQSDTR
TVRAILEYQADKVALDANQTPAESASHN"

complement(1416..1907)

/locus_tag="RB2939"

complement(1416..1907)

/locus_tag="RB2939"

/note="PMID: 8943436 best DB hits: BLAST: embi:CAB66271.1;

(AL136519) putative integral-membrane protein.; E=1e-06
emb1:CAB72202.1; (AL138851) putative integral-membrane
protein; E=5e-06"

/codon_start=1

/transl_table=11

/product="probable integral-membrane protein"

/protein_id="CAD72911.1"

/db_xref="GI:32443469"

/translation="MFFHTVLVSATLLCSLVAGFTFAFDVVAMPGRUNDRSFRFAF
QVIGVINDQNFPMFVWGAIAIVAAVIGFSQIDGVDRTLLIVAAVYLLGVQLP
TVIRINPLNNKIQAIDVEQDKDQDELLAARVFEAQWNRNVRVTYVTSITIALSHVLL
LRI"

1930..2370

/locus_tag="RB2940"

1930..2370

/locus_tag="RB2940"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CAD72912.1"

/db_xref="GI:32443470"

/translation="MTAASVRLILHSTTGMGCLGNVACGHRDVAVRLVIGYRGES
EGDCEKIDPHEDHGSKRLVAAVSLVQHYGKTSYWELICETEPKGFHVDAYFCVSLIV
PLSHAWVRSGDGRSTSCNVTSPICVRRATATRELLSHETLRPHQ"

complement(2321..2797)

/locus_tag="RB2941"

complement(2321..2797)

/locus_tag="RB2941"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CAD72913.1"

/db_xref="GI:32443471"

/translation="MSIVSLVTAISIGAPFQISDTENVLEREDAGIQIQGRPEAGT
GLCGASLRVPSDVYVSLPPRGENDAHRLRLRRPPLHIQIKRGPVTEAEHELLES
NVTGAMLLYDIRLDDGNLKKRRSPROGFRFADEDVNGDGIIDEDDGEIRDSANGG"

2775..3035

/locus_tag="RB2943"

2775..3035

/locus_tag="RB2943"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CAD72914.1"

/db_xref="GI:32443472"

/translation="MTSHTIDQQSFVAVHSNIPDRGGEVEFFPHGKPCELSSDG
CPPUCYQIGSAAKTLTARTQFAFGQSQSVHLLAFRN"

3042..3182

/locus_tag="RB2945"

3042..3182

/locus_tag="RB2945"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CAD72915.1"

/db_xref="GI:32443473"

/translation="MASQTVTGEQGVLTSTCGETCKSCRRLLAEGNAKILSIKQV
GS"

3188..3334

/locus_tag="RB2946"

3188..3334

/locus_tag="RB2946"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CAD72916.1"

/db_xref="GI:32443474"

/translation="MRGFAANRAAASIAVSIQQSRPNYALRSRELLGRABQORLSS
KCAP"

complement(3403..3849)

/locus_tag="RB2948"

complement(3403..3849)

/locus_tag="RB2948"


```
/codon_start=1
/transl_table=11

Query Match      7.4%; Score 105.6; DB 1; Length 321950;
Best Local Similarity 50.9%; Pred. No. 8e-16;
Matches 330; Conservative 0; Mismatches 309; Indels 9; Gaps 3;

QY      6  CTTCCCAAGAGAACTTCAGGCTCTTGGGTGCAACCATCACGCCAATGACTGAGAA 65
Db      205761 CATGGCAGGCGCAAACTATCGGCGCTGATCGCGCAACCTACACCCGATGAAGCGTG 205702

QY      66  TGGAGAAATCAACTTTTCAGTAAATGGTCAATATGGATATGATATCTTGTGAAGAAACAGGG 125
Db      205701 CGCGCATCTCAACCT---CGAATGTTTCCCGCCCATGGTGGAGAGTTGCTACTTGAACGG 205645

QY      126  AGTGAAGAACATTTTGTGAATGGCAACACAGAGAGAGGCTGTCCCTGAGCGCTCTCAGA 185
Db      205644  AGTCTCAGGACTGTAGTCTCGCGCAGCACCGCGGAGGAAATGTGCTCACACGCGAGA 205585

QY      186  GCCTCGCCAGGTTGCAAGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAGGTGAT 245
Db      205584  ACCCCAGCTGGTCGCTCGCTTTCGTCA---ATGCGACGCGCGCGAGTTCCTCGTGTAT 205528

QY      246  AATTCAGTAGGAGCACTGAGCTTGAGGAGTCAACAGGAACTGGCCCAACATCGACGAGA 305
Db      205527  CGTTCAGTTGGGCACAAACAGTTTGGGTGAGGACACAGAACTGGCCCAACATGCCACGCA 205468

QY      306  AATAGAGCTGATGGCATCGCTGTCATTTGCAACCGTTCCTCTCAAGCCATGGACCAAGA 365
Db      205467  GATCGGTGCCAGTGCAGATTTCGGGACCTGCCCATCGTATTTCAAAGTTGCCAGCGTCA 205408

QY      366  TATCCTGATTAAATTCCTAAAGGAGTGGCTGTGCGCGCCCTGCGCCCTGCCATTTATTA 425
Db      205407  GTCGCTCACCTCTTTGCATGCAAGAACTAGCGGCGCGCACCGGAGACGCGCTTTACTA 205348

QY      426  CTATCACATTCCTCCCTTGACAGGGGTAAAGATTCTGCTGAGGAGTTGTTGGATGGGAT 485
Db      205347  TTACCAACATCCCGGTGCTAACGGGTTCCAGCATCGACATGGTGGAAATCTTGACTCAGC 205288

QY      486  TCTGGATAAGATCCCACTTCCAGGCGCTGAATTCAGTGATACAGATCTCTTAGACTT 545
Db      205287  GGCAGAGCCATCCCGACTTTGGTGGACTGAATATACCGATACGATGTTGTTGAGTT 205228

QY      546  CGGGCAATGTGTTGATCAGAAATCCCAAGCAACAGTTTCTTTCTTTTGGGSGGATGA 605
Db      205227  TCAACGGTGCCTCGAATGTGGAATCGGA---AGTTTGAAGTCTTTGGGSGCTGCGACGA 205171

QY      606  GCAACTGTTGAGTCTCTGTTGATGGAGCAACTGGAGCAGTGGCGAG 653
Db      205170  AATGTTATTGGGTGCGACCGCGAGCGGTGCCGTGCCGATTGGAAG 205123
```

Search completed: May 11, 2004, 17:15:08
Cdb time : 5814.58 secs

GenCore version: 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2004, 04:14:30 ; Search time 616.115 Seconds
(without alignments)
9853.151 Million cell updates/sec

Title: US-09-930-440B-1

Perfect score: 1429

Sequence: 1 atggccttcccaagaagaa.....cttttgatcatttogatg 1429

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 3%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1429	100.0	1429	3	AA50567 Human ald
2	1429	100.0	1429	3	AA337764 Human gly
3	1429	100.0	1429	4	Aah28456 Nucleotid
4	1429	100.0	1429	7	Aal53991 DNA encod
5	1332.4	93.2	2536	2	Aaz22092 cDNA enco
6	1332.4	93.2	2536	3	Aaz298699 Human N-a
7	1090.8	76.3	2441	6	Adn86096 Polyadeny
8	668.2	46.8	1171	4	AA193000 Human pol
9	652.4	45.7	1483	5	Aha52191 Human dih
10	433	30.3	444	5	Aha52191 Human APP
11	352	24.6	753	4	Aak63371 Human imm
12	128.2	9.0	472	8	Ach21667 Human adu
13	124	8.7	628	7	Az219274 Group III
14	95.4	6.7	14904	4	AAK75386 Human imm
15	77.2	5.4	618	5	ABA09731 Human bon
16	69.8	4.9	4646	6	Aak68979 Human imm
17	53.4	3.7	912	6	Abn66507 Streptoco
18	53	3.7	29255	4	Aas59516 Propionib
19	53	3.7	29255	7	ACf64445 Propionib
20	48.4	3.4	8056	7	Abz10246 Haematopo
21	46.8	3.3	2000	7	Ada71938 Rice gene
22	45.4	3.2	6117	6	Ab133025 Human imm
23	44.8	3.1	5771	6	Ab133951 Human imm

24	44.4	3.1	6109	6	ABL32326	Ab132326 Human imm
25	44.4	3.1	6109	6	AA561077	AA561077 Human gen
26	44.2	3.1	1215	2	AAT58339	Aat58339 Tobacco e
27	44.2	3.1	4253	2	AAT58361	Aat58361 Tobacco e
28	44.2	3.1	19380	6	AA561426	AA561426 Human gen
29	44	3.1	8170	6	ABK28257	ABK28257 DNA trans
30	44	3.1	13574	6	ABL33317	ABL33317 Human imm
31	43.8	3.1	2000	7	ADA71938	Ada71938 Rice gene
32	43.4	3.0	915	6	ABN66506	Abn66506 Streptoco
33	43.4	3.0	110000	6	ABN71527	Abn71527 Streptoco
34	43.2	3.0	915	7	ABX07056	Abx07056 S. pneumo
35	43.2	3.0	930	7	ABZ42256	Abz42256 Streptoco
36	43.2	3.0	6361	6	ABL33141	Ab133141 Human imm
37	43.2	3.0	9897	2	AAV52143	AAV52143 Streptoco
38	43.2	3.0	17848	4	AA545323	AA545323 Chemical
39	43.2	3.0	17848	6	ABK39976	Abk39976 Human che
40	43.2	3.0	17848	6	ABK28164	ABK28164 DNA trans
41	43.2	3.0	110000	7	AB556454	Continuation 12 o
42	42.2	3.0	1075	2	AA84335	Aax84335 Stealth v
43	42.2	3.0	8991	6	ABK31393	Abk31393 Signal tr
44	42.2	3.0	8991	6	ABL70340	Ab170340 Chemical
45	42.2	3.0	8991	6	AA561295	AA561295 Human gen

ALIGNMENTS

RESULT 1
AAA50567
ID AAA50567 standard; cDNA; 1429 BP.

AC AAA50567;

DT 19-DEC-2000 (first entry)

DE Human aldolase cDNA.

KW Aldolase; human; sialylation; glycoprotein; plasminogen; transferrin;

KW thyrotropin; Na+,K+-ATPase; ss.

OS Homo sapiens.

EH Key Location/Qualifiers

FT CDS 1..693

FT /*tag= a

FN WO200052135-A2.

PD 08-SEP-2000.

PF 01-MAR-2000; 2000WO-US005313.

PR 02-MAR-1999; 99US-0122582P.

PR 08-DEC-1999; 99US-0169624P.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (UYJO) UNIV JOHNS HOPKINS.

PA (UYWY-) UNIV WYOMING.

DR Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;

DR WPI; 2000-572178/53.

DR P-FSDB; AAY96099.

Recombinant production of sialylated glycoproteins using cells in which the expression of enzymes, e.g. sialic acid synthetase, involved in the sialylation reaction has been altered.

XX Disclosure; Page 98-99; 144pp; English.

PS The present sequence is that of human cDNA encoding aldolase (see

CC AAY96099). The invention provides methods and recombinantly engineered

CC cells for producing glycoproteins having sialylated oligosaccharides. The

CC methods involve altering the expression of enzymes involved in
 CC carbohydrate processing. A claimed cell producing sialylated glycoprotein
 CC above endogenous levels expresses at least 1 enzyme selected from GlcNAc-
 CC 2 epimerase, an enzyme catalyzing the conversion of UDP-GlcNAc to ManNAc,
 CC sialic acid synthetase, aldolase, CMP-sialic acid synthetase and CMP-
 CC sialic acid transporter at above endogenous levels. Endogenous N-
 CC acetylglucosaminidase activity may be suppressed. A claimed method for
 CC manipulating glycoprotein in an insect cell comprises enhancing the
 CC expression of 1 of the above enzymes, and a claimed method for producing
 CC sialylated glycoproteins involves expressing a heterologous protein
 CC (especially plasmidogen, transferrin, Na⁺ K⁺-ATPase or thyrotropin) in
 CC the insect cell. Yeast, insect, fungal, plant and bacterial host cells
 CC can be engineered to produce new forms of sialylated glycoproteins,
 CC higher concentrations of sialylated glycoproteins and/or elevated
 CC concentrations of donor substrates (e.g. nucleotide sugars) required for
 CC sialylation
 CC
 XX
 SQ Sequence 1429 BP; 400 A; 284 C; 324 G; 421 T; 0 U; 0 Other;

Query Match 100.0%; Score 1429; DE 3; Length 1429;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCTTCCCAAGAGAACTTCAGGCTCTTGCTGCTGACACCTACGCGCATGACT 60
 DB 1 ATGGGCTTCCCAAGAGAACTTCAGGCTCTTGCTGCTGACACCTACGCGCATGACT 60
 QY 61 GAGATGGAGAAATCAACTTTTCAGTAATGCTGATGATGATGATGATGATGATGAT 120
 DB 61 GAGATGGAGAAATCAACTTTTCAGTAATGCTGATGATGATGATGATGATGATGAT 120
 QY 121 CAGGAGTCAAGCAATTTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 180
 DB 121 CAGGAGTCAAGCAATTTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 180
 QY 181 TCAGAGCTGCGCAGGAGTGGGTGACAAAGAGGAGGACAAAGCTGATGATGATGATGAT 240
 DB 181 TCAGAGCTGCGCAGGAGTGGGTGACAAAGAGGAGGACAAAGCTGATGATGATGATGAT 240
 QY 241 GTGATAATTCAGTGAAGGACATGAGCTTGAAGGAGTCAAGGAGTGGCCCAAGTCA 300
 DB 241 GTGATAATTCAGTGAAGGACATGAGCTTGAAGGAGTCAAGGAGTGGCCCAAGTCA 300
 QY 301 GCAGAAATAGGAGCTGATGGCACTGCTGCTCATGACAGCTTCTTCTCAAGCCATGGACC 360
 DB 301 GCAGAAATAGGAGCTGATGGCACTGCTGCTCATGACAGCTTCTTCTCAAGCCATGGACC 360
 QY 361 AAAGATATCTGATTAATTTCTTAAAGGAGTGGGTGCTGCTGCGCGCCCTGCCCTGCCATTT 420
 DB 361 AAAGATATCTGATTAATTTCTTAAAGGAGTGGGTGCTGCTGCGCGCCCTGCCCTGCCATTT 420
 QY 421 TATTAATATCAATTCCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 480
 DB 421 TATTAATATCAATTCCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 480
 QY 481 GGGATCTGGATTAAGTATCCCACTCCAGGCTGGAATTCAGTGATACAGATCTCTTA 540
 DB 481 GGGATCTGGATTAAGTATCCCACTCCAGGCTGGAATTCAGTGATACAGATCTCTTA 540
 QY 541 GACTTCGGGCAATGTTGATCAGAAATGCCAGCAACAGTTTGTCTTTTGGGGTG 600
 DB 541 GACTTCGGGCAATGTTGATCAGAAATGCCAGCAACAGTTTGTCTTTTGGGGTG 600
 QY 601 GATGAGCAACTGTTGATGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 660
 DB 601 GATGAGCAACTGTTGATGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 660
 QY 661 TCCAGAGATTTATCACTTTTGTGTAACATGATGATGATGATGATGATGATGATGATGATGAT 720
 DB 661 TCCAGAGATTTATCACTTTTGTGTAACATGATGATGATGATGATGATGATGATGATGATGAT 720
 QY 721 TCATGACTCTGGTCTCTGGGATTCCTAATGGGCGCCACCGCGCTTCCATGCGAGAAAGCT 780

DB 721 TCATGACTCTGGTCTCTGGGATTCCTAATGGGCGCCACCGCGCTTCCATGCGAGAAAGCT 780
 QY 781 CCAGGAGATTTACTGATAGTGTGAAGCTAAACTGAAGAGCTTGGATTTCTTTCTTTCA 840
 DB 781 CCAGGAGATTTACTGATAGTGTGAAGCTAAACTGAAGAGCTTGGATTTCTTTCTTTCA 840
 QY 841 CTGATTTAAGAGTGAAGAACTTGAAGCTGGTGTAGTGTAGTGTCTCTATCAAACTCAGGT 900
 DB 841 CTGATTTAAGAGTGAAGAACTTGAAGCTGGTGTAGTGTAGTGTCTCTATCAAACTCAGGT 900
 QY 901 TTGCACCTTGAGACATTAATCTACCTTAATAGTGTGATTTTCTCAGGAAATTTAGAT 960
 DB 901 TTGCACCTTGAGACATTAATCTACCTTAATAGTGTGATTTTCTCAGGAAATTTAGAT 960
 QY 961 GAACTTGAATAAATCTCTCTAGCAAAATGAATCTCAAAATGAAGCTTGAAGTACCTTTTG 1020
 DB 961 GAACTTGAATAAATCTCTCTAGCAAAATGAATCTCAAAATGAAGCTTGAAGTACCTTTTG 1020
 QY 1021 TGAGCCTTAAAAGTCTTATTTTGTGAAGGGCAAAAATCTTAGGAGTCAAACTCTCAG 1080
 DB 1021 TGAGCCTTAAAAGTCTTATTTTGTGAAGGGCAAAAATCTTAGGAGTCAAACTCTCAG 1080
 QY 1081 TCATTCATTTCAAGATTTTCTGAGGAAATTTCTGTTATATGATGAAATGGAATC 1140
 DB 1081 TCATTCATTTCAAGATTTTCTGAGGAAATTTCTGTTATATGATGAAATGGAATC 1140
 QY 1141 AAGAGGAAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200
 DB 1141 AAGAGGAAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200
 QY 1201 TCTGTTCTTAATCTTATTTTAAAGTGTCTAATTTTAAACCACTATATATGCTTCAT 1260
 DB 1201 TCTGTTCTTAATCTTATTTTAAAGTGTCTAATTTTAAACCACTATATATGCTTCAT 1260
 QY 1261 TTTAATAAATTTCAATTTGGAATCTAGGAAATCTGAGCTACTGCTATTTAGGAGGAC 1320
 DB 1261 TTTAATAAATTTCAATTTGGAATCTAGGAAATCTGAGCTACTGCTATTTAGGAGGAC 1320
 QY 1321 TTTAATAAATTTCAATTTGGAATCTGCTCACTGATATCAACTCAAAATPACACAGCTCAT 1380
 DB 1321 TTTAATAAATTTCAATTTGGAATCTGCTCACTGATATCAACTCAAAATPACACAGCTCAT 1380
 QY 1381 GGCTGCTCAGTCTAATCTAGAAATGAGATGATGATGATGATGATGATGATGATGATGATG 1429
 DB 1381 GGCTGCTCAGTCTAATCTAGAAATGAGATGATGATGATGATGATGATGATGATGATGATG 1429

RESULT 2
 AAA37764

ID AAA37764 standard; DNA; 1429 BP.

XX AAA37764;

XX 04-DEC-2000 (first entry)

XX Human glycosylation enzyme clone HDPK85 coding sequence.

DE Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
 KW immunotherapy; cosmetic surgery; metabolism; immune system disorder;
 KW haematopoietic cell deficiency; blood coagulation disorder; asthma;
 KW afibrinogenemia; blood platelet disorder; thrombocytopaenia; neoplasia;
 KW autoimmune disorder; Addison's disease; multiple sclerosis; purpura;
 KW allergic encephalomyelitis; allergic reaction; organ rejection;
 KW graft-versus-host disease; inflammation; hyperproliferative disorder;
 KW sarcoidosis; infection; gene therapy; CMP sialic acid synthetase; ds.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH 1. .593
 FT CDS /*tag= a
 FT /product= "aldolase"

PN WO200052:36-A2.
 XX 08-SEP-2000.
 XX 01-MAR-2000; 2000WC-US005325.
 XX 02-MAR-1999; 99US-0122409P.
 XX (HUYA-) HUMAN GENOME SCI INC.
 PA Co-eman TA;
 XX WPI: 2000-572179/53.
 DR P-PSDB; AAY90353.
 XX
 PT New human glycosylation enzymes cytidine 5'-monophosphate sialic acid
 PT synthetase, sialic acid synthetase and aldolase and nucleic acids
 PT encoding the proteins for treating e.g., immune system disorders,
 PT microbial diseases.
 XX
 PS Claim 4; Page 113-114; 115pp; English.
 CC This sequence encodes a human glycosylation enzyme clone of the
 CC invention, designated HDPK85. The protein of this clone is an aldolase.
 CC The sequences are useful as reagents for the differential identification
 CC of the tissues or cell types present in a biological sample, as
 CC immunological probes, for treating a disease or condition resulting from
 CC under expression of such polypeptide, for the detection and/or treatment
 CC of disorders involving aberrant glycolysis, e.g. cramps, myoglobinuria,
 CC and as tumour marker and/or immunotherapy targets. They may also be used
 CC to differentiate, proliferate and attract cells leading to the
 CC regeneration of tissues, to modulate mammalian characteristics (e.g. in
 CC cosmetic surgery) or mammalian metabolism affecting catabolism, anabolism
 CC processing, utilisation and energy storage, to change a mammal's mental
 CC state by influencing and as a food additive or preservative. The proteins
 CC can be used to assay protein levels in a sample, as a marker or detector
 CC of an immune system disorder, to inhibit cytokine activity, and as a
 CC vaccine. They may further be used to treat immune system or of
 CC haematopoietic cell deficiencies or disorders, blood coagulation
 CC disorders (e.g. afibrinogenemia), blood platelet disorders (e.g.
 CC thrombocytopaenia), wounds resulting from trauma or surgery, autoimmune
 CC disorders (e.g. Addison's disease, multiple sclerosis, allergic
 CC encephalomyelitis), allergic reactions (e.g. asthma), organ rejection,
 CC graft-versus-host disease, inflammation, hyperproliferative disorders
 CC (e.g. neoplasia, purpura, sarcoidosis), diseases caused by viruses (e.g.
 CC hepatitis, meningitis, AIDS), bacteria and fungi (including e.g.
 CC tuberculosis, conjunctivitis, sepsis, typhoid, chlamydia, cellulitis),
 CC and diseases caused by parasites (e.g. amoebiasis, coccidiosis,
 CC leishmaniasis, scabies, malaria, toxoplasmosis).
 XX
 SQ Sequence 1429 BP; 400 A; 284 C; 324 G; 421 T; 0 U; 0 Other;

241 GTGATAATTCAGTATGAGACCTGAGCTTGAAGAGTCAAGGAACTGGCCCAACATGCA 300
 301 GCAGAAATAGAGAGCTGATGGCATCGCTGTCATTGACACGCTTCTTCTCAACCCATGGACC 360
 301 GCAGAAATAGAGAGCTGATGGCATCGCTGTCATTGACACGCTTCTTCTCAACCCATGGACC 360
 361 AAGAGATATCTGATTAATTTCTTAAAGAGAGTGGCTGCTCCGCCCTGCCCTCCCATTT 420
 361 AAGAGATATCTGATTAATTTCTTAAAGAGAGTGGCTGCTCCGCCCTGCCCTCCCATTT 420
 421 TATTACTATCACATTTCTGCTTGCAGAGGGTAAAGATTCTGCTGAGGAGTGTGGAT 480
 421 TATTACTATCACATTTCTGCTTGCAGAGGGTAAAGATTCTGCTGAGGAGTGTGGAT 480
 481 GGGATTTGATTAAGATCCCAACCTTCCAAAGGGGTGMAATTCAGTGATACAGATCTCTTA 540
 481 GGGATTTGATTAAGATCCCAACCTTCCAAAGGGGTGMAATTCAGTGATACAGATCTCTTA 540
 541 GACTTGGGGCAATGTCTGATCAGATCGCCAGCAACAGTGTGCTTCTTCTTTGGGGTG 600
 541 GACTTGGGGCAATGTCTGATCAGATCGCCAGCAACAGTGTGCTTCTTCTTTGGGGTG 600
 601 GATGAGCAACTGTGAGTGTCTGCTGATGGAGCAACTGGAGCACTGGGCGAGTTTGTGA 660
 601 GATGAGCAACTGTGAGTGTCTGCTGATGGAGCAACTGGAGCACTGGGCGAGTTTGTGA 660
 661 TCCAGAGATTTATCAACTTTTGTGTCAAACTAGGTTTGGAGTGTCAAGACCAAGCCA 720
 661 TCCAGAGATTTATCAACTTTTGTGTCAAACTAGGTTTGGAGTGTCAAGACCAAGCCA 720
 721 TCATGACTGTGCTCTGAGTTCATGAGTGGCCAGCCAGCCGCTTCCACTGCAGAAAGCCT 780
 721 TCATGACTGTGCTCTGAGTTCATGAGTGGCCAGCCAGCCGCTTCCACTGCAGAAAGCCT 780
 781 CCAGGAGTTTACTGATAGTCTGAAGCTAAACTGMAAGAGCTTGGATTTCTTCTTTTCA 840
 781 CCAGGAGTTTACTGATAGTCTGAAGCTAAACTGMAAGAGCTTGGATTTCTTCTTTTCA 840
 841 CTGATTTAAAGATGGAACCTTGAAGCTGATAGTGGCTCTCTATCAAAATCAGGGT 900
 841 CTGATTTAAAGATGGAACCTTGAAGCTGATAGTGGCTCTCTATCAAAATCAGGGT 900
 901 TTCACCTTTCAGACATATCTACTTAATAGTGCCTTTTCTCTAGGAGTATTTAGAT 960
 901 TTCACCTTTCAGACATATCTACTTAATAGTGCCTTTTCTCTAGGAGTATTTAGAT 960
 961 GAACCTTGAATAAACTCTCTCAGCAAAATGAAATCTCACAATAAGCATTCAGGTACCTTTG 1020
 961 GAACCTTGAATAAACTCTCTCAGCAAAATGAAATCTCACAATAAGCATTCAGGTACCTTTG 1020
 1021 TGAGCCTTAAAGATCTTTTGTGAGGGGCAAAACTCTAGGAGTCAAACTCTCAG 1080
 1021 TGAGCCTTAAAGATCTTTTGTGAGGGGCAAAACTCTAGGAGTCAAACTCTCAG 1080
 1081 TCATTCATTTTCAGATTTTGTGAGGAAATTTCTGTTATATGATGAATGGAATC 1140
 1081 TCATTCATTTTCAGATTTTGTGAGGAAATTTCTGTTATATGATGAATGGAATC 1140
 1141 AAGAGGAAATTTGATTTGATTAATTCATCTGCTTTTAGGAGCTCTCATTTCTCGCTC 1200
 1141 AAGAGGAAATTTGATTTGATTAATTCATCTGCTTTTAGGAGCTCTCATTTCTCGCTC 1200
 1201 TCTGGTCTTAAATCTTAAAGTGTCTAATTTTAAACCACTATATATATGCTCTCAT 1260
 1201 TCTGGTCTTAAATCTTAAAGTGTCTAATTTTAAACCACTATATATATGCTCTCAT 1260
 1261 TTTAAATAATTAATTAATTTGGAATCTAGGAAACTCTGAGTCTGCTATTTAGGAGGAC 1320
 1261 TTTAAATAATTAATTAATTTGGAATCTAGGAAACTCTGAGTCTGCTATTTAGGAGGAC 1320
 1321 TTTAAATCAAACTGTAAACATGTCTCAACTGTATACAACTCAAAATACAGAGCTCATTT 1380

Query Match 100.0%; Score 1429; DB 3; Length 1429;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 1429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGCTTCCCAAGAGAACTTCAGGCTTGTGCTGCAACCATCAGCCCAATGACT 60
 DB 1 ATGGCTTCCCAAGAGAACTTCAGGCTTGTGCTGCAACCATCAGCCCAATGACT 60
 QY 61 GAGATGGAGAAATCACTTTTCAGTAATGCTGCTGATGCTGATGCTGTAAGAA 120
 DB 61 GAGATGGAGAAATCACTTTTCAGTAATGCTGCTGATGCTGTAAGAA 120
 QY 121 CAGGAGTGAAGAACTTTTGTGTAATGTCACACAGGAGAGCCCTGCTCAGCGTC 180
 DB 121 CAGGAGTGAAGAACTTTTGTGTAATGTCACACAGGAGAGCCCTGCTCAGCGTC 180
 QY 181 TCAGAGCTGCCAGGTTGAGAGGAGTGGTGCACAAAGGAGAGCAGCTGGATCAG 240
 DB 181 TCAGAGCTGCCAGGTTGAGAGGAGTGGTGCACAAAGGAGAGCAGCTGGATCAG 240
 QY 241 GTGATAATTCAGTATGAGACCTGAGCTTGAAGAGTCAAGGAACTGGGCCCAACATGCA 300

DB 1321 TTTATACCAACTCTAAGTCTCTCAACTGTATCAACTCAACTCAAAATACACAGCTCATTT 1380

QY 1381 GGGCTGCTCAGTCTAAGTCTCTAAGTCTCTCAACTGTATCAACTCAACTCAAAATACACAGCTCATTT 1429

DB 1391 GGGCTGCTCAGTCTAAGTCTCTAAGTCTCTCAACTGTATCAACTCAACTCAAAATACACAGCTCATTT 1429

RESULT 3

AAH28456

ID AAH28456 standard; DNA; 1429 BP.

XX AC AAH28456;

XX 17-SEP-2001 (first entry)

XX Nucleotide sequence of a human aldolase.

XX Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;

XX cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;

XX sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;

XX vaccine; ss.

XX Homo sapiens.

XX Key location/Qualifiers

XX CDS 1..693

XX /tag= a

XX /product= "aldolase"

XX WO200142492-A1.

XX 14-JUN-2001.

XX 07-DEC-2000; 2000WO-US033136.

XX 09-DEC-1999; 99US-0169839P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (UYJO) UNIV JOHNS HOPKINS.

XX (UYEM) UNIV TEMPLE.

XX (UYWY-) UNIV WYOMING.

XX Berenbaugh XJ, Lawrence S, Lee YC, Coleman TA, Palter K;

XX Jarvis D;

XX WPI; 2001-44-575/47.

XX P-PSDB; AAB84681.

XX Cells producing cytidine monophosphate-sialic acid and sialylated

XX glycoprotein above endogenous levels for production of vaccines and

XX therapeutics.

XX Disclosure; Fig 27; 182pp; English.

XX The specification describes a method for manipulating carbohydrate

XX processing pathways in cells of interest. The methods are used to

XX manipulate multiple pathways involved with the sialylation reaction by

XX using recombinant DNA technology and substrate feeding approaches to

XX enable the production of sialylated glycoproteins in the cells. The

XX sialylation process involves the post-translational addition of the donor

XX substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific

XX acceptor carbohydrate. The cells express at least one enzyme, selected

XX from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,

XX CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The

XX cells are useful for producing complex sialylated glycoproteins in cells

XX of interest, especially insect cells. Glycoproteins containing sialylated

XX oligosaccharides are useful as vaccines, therapeutics and diagnostic

XX tools. Cells producing complex sialylated glycoproteins are useful for

XX enhancing the value of heterologous expression systems and increasing the

XX application of heterologous cell expression products as vaccines,

XX therapeutics and diagnostic tools as well as increasing the variety of

XX heterologous proteins that can be produced and lowering biotechnology

XX production costs. The present sequence encodes a human aldolase, which is

CC used in the method of the invention

XX Sequence 1429 BP; 400 A; 284 C; 324 G; 421 T; 0 U; 0 Other;

QY Query Match 100.0%; Score 1429; DB 4; Length 1429;

DB Best Local Similarity 100.0%; Pred. No. 0;

Matches 1429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCTTCCCAAGAGAAACTTCAGGGTCTTGTGGCTGGTGCACCAACCATCAGCGCATGACT 60

DB |||||||

QY 1 ATGGCCTTCCCAAGAGAAACTTCAGGGTCTTGTGGCTGGTGCACCAACCATCAGCGCATGACT 60

DB |||||||

QY 61 GGAATGGAGAAATCAACTTTTCAGTAAATGGTGCAGTATGTGATTTATCTTGTGAAGAA 120

DB |||||||

QY 61 GGAATGGAGAAATCAACTTTTCAGTAAATGGTGCAGTATGTGATTTATCTTGTGAAGAA 120

DB |||||||

QY 121 CAGGAGTCAAGACATTTTGTGAATGGCCACACAGGAGAGAGGCTGTCCCTGAGCGTC 180

DB |||||||

QY 121 CAGGAGTCAAGACATTTTGTGAATGGCCACACAGGAGAGAGGCTGTCCCTGAGCGTC 180

DB |||||||

QY 181 TCAGAGCGTCCGACAGGTTGCAGAGGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAG 240

DB |||||||

QY 181 TCAGAGCGTCCGACAGGTTGCAGAGGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAG 240

DB |||||||

QY 241 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGGAGTGCACAGGAACCTGCCCAACATGCA 300

DB |||||||

QY 241 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGGAGTGCACAGGAACCTGCCCAACATGCA 300

DB |||||||

QY 301 GCAGAAATAGGAGCTGATGGCATCGCTTCANTGCAACCGTTCTTCTCAAGCCATGGACC 360

DB |||||||

QY 301 GCAGAAATAGGAGCTGATGGCATCGCTTCANTGCAACCGTTCTTCTCAAGCCATGGACC 360

DB |||||||

QY 361 AAAGATATCTGATTAATTTCTTAAGGAGTGGCTGTCCGCCCTCCCTGCGCATTT 420

DB |||||||

QY 361 AAAGATATCTGATTAATTTCTTAAGGAGTGGCTGTCCGCCCTCCCTGCGCATTT 420

DB |||||||

QY 421 TATTACTATCACTTCTGCTTGCAGGGTAAAGATTCTGCTGAGGAGTGTGGAT 480

DB |||||||

QY 421 TATTACTATCACTTCTGCTTGCAGGGTAAAGATTCTGCTGAGGAGTGTGGAT 480

DB |||||||

QY 481 GGGATTCTGGATPAAGATCCCACTTCCAAAGGCTGAAATTCAGTGATACAGATCTCTTA 540

DB |||||||

QY 481 GGGATTCTGGATPAAGATCCCACTTCCAAAGGCTGAAATTCAGTGATACAGATCTCTTA 540

DB |||||||

QY 541 GACTTCGGGCAATGTGTGATCGAATGCCAGCAACAGTTTCTTCTTTTGGGGTG 600

DB |||||||

QY 541 GACTTCGGGCAATGTGTGATCGAATGCCAGCAACAGTTTCTTCTTTTGGGGTG 600

DB |||||||

QY 601 GATGAGCAACTGTTGAGTGTCTGCTGATGGAGCAACTGGAGCACTGGGCACTTTTGA 660

DB |||||||

QY 601 GATGAGCAACTGTTGAGTGTCTGCTGATGGAGCAACTGGAGCACTGGGCACTTTTGA 660

DB |||||||

QY 661 TCCAGAGATTTATCAACTTTGTCTCAACTAGGTTTGGAGTGTGCACAGCAACAGCCA 720

DB |||||||

QY 661 TCCAGAGATTTATCAACTTTGTCTCAACTAGGTTTGGAGTGTGCACAGCAACAGCCA 720

DB |||||||

QY 721 TCATGACTGTGCTCTCTGCTGATCCCAAGGCTCCAGCTTCCACTGCAGAAAGCT 780

DB |||||||

QY 721 TCATGACTGTGCTCTCTGCTGATCCCAAGGCTCCAGCTTCCACTGCAGAAAGCT 780

DB |||||||

QY 781 CAGGAGTCTTACTGATAGTGTGAAGCTAAACTGAAGAGCTGGATTTCTTTCTTCA 840

DB |||||||

QY 781 CAGGAGTCTTACTGATAGTGTGAAGCTAAACTGAAGAGCTGGATTTCTTTCTTCA 840

DB |||||||

QY 841 CTGATTTAAGATGGAAACTTGGAGCTGTAGTGTGCTCTCTATCAATCAGGT 900

DB |||||||

QY 841 CTGATTTAAGATGGAAACTTGGAGCTGTAGTGTGCTCTCTATCAATCAGGT 900

DB |||||||

QY 901 TTGACCTTGGACATATCTACCTTAAATAGTGCATTTTCTCAGGAAATTTAGAT 960

DB |||||||

QY 901 TTGACCTTGGACATATCTACCTTAAATAGTGCATTTTCTCAGGAAATTTAGAT 960

DB |||||||

QY 961 GAACTTGAATAAACTCTCTAGCAAAATGAAATCTCAAAATAGCAATGAGGTACCTTTTG 1020

Db 961 GAACITGAATAAATCTCTAGCAAAATGAATCTCAATAAAGCATGTAGGTACCTTTTG 1020
QY 1021 TGAGCCTTAAAGAGTCTTATTTTGTGAAGGGGCAAAABCTTAGGAGTCAAACTCTCAG 1080
Db 1021 TGAGCCTTAAAGAGTCTTATTTTGTGAAGGGGCAAAABCTTAGGAGTCAAACTCTCAG 1080
QY 1081 TCATTCATTTACAGATTTTTTTTGTGAGAGAAATCTCTGTATATATGATGAATGGAATC 1140
Db 1081 TCATTCATTTACAGATTTTTTTTGTGAGAGAAATCTCTGTATATATGATGAATGGAATC 1140
QY 1141 AAGAGAGAAATTTGAATTAATTTCCATCTGCTTTTAGAGCTCTCATATCTCGGTC 1200
Db 1141 AAGAGAGAAATTTGAATTAATTTCCATCTGCTTTTAGAGCTCTCATATCTCGGTC 1200
QY 1201 TCTGGTTCCTAATCCATTTTAAAGTTGTCTAATTTTAAACCACTATAATATGTCAT 1260
Db 1201 TCTGGTTCCTAATCCATTTTAAAGTTGTCTAATTTTAAACCACTATAATATGTCAT 1260
QY 1261 TTTAATAAATATTCATTTTGAATCTAGGAAACTCTGAGCTACTGCTATTTAGGAGGCAC 1320
Db 1261 TTTAATAAATATTCATTTTGAATCTAGGAAACTCTGAGCTACTGCTATTTAGGAGGCAC 1320
QY 1321 TTTAATACCAAACTGTAACATGCTCAACTGTATAGAACTCAAAATACACCAGTCAATTT 1380
Db 1321 TTTAATACCAAACTGTAACATGCTCAACTGTATAGAACTCAAAATACACCAGTCAATTT 1380
QY 1381 GGCTGCTCAGTCTAATCTAGAAATGGATGCTTTTGAATTCATTTTGGATG 1429
Db 1381 GGCTGCTCAGTCTAATCTAGAAATGGATGCTTTTGAATTCATTTTGGATG 1429

RESULT 4
AAL53991
ID AAL53991 standard; DNA; 1429 BP.

XX AAL53991;

DT 06-MAR-2003 (first entry)

DE DNA encoding a human aldolase protein.

XX Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;
KW GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)Nac; sialic acid;
KW synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;
KW transporter; sialylated glycoprotein; human; aldolase; gene; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
FT I. .693
CDS /tag= a
/product= "Human aldolase protein"

US2002142386-A1.

FT 03-OCT-2002.

XX 16-AUG-2001; 2001US-00930440.

XX 02-MAR-1999; 99US-0122582P.

XX 08-DEC-1999; 99US-0169624P.

XX 25-AUG-2000; 2000US-0227579P.

XX (BETE/) BETENBAUGH M J.

XX (LAWR/) LAWRENCE S.

XX (LEEV/) LEE Y C.

XX (COLE/) COLEMAN T A.

XX PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;

XX WPI; 2003-102519/09.

DR P-PSDB; AA026544.

XX Manipulating glycoprotein production in insect cell, involves enhancing
PT expression of enzymes involved in carbohydrate processing pathway such as
PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.
XX Disclosure; Fig 27; 88pp; English.

XX The invention relates to a novel method for manipulating glycoprotein
CC production in an insect cell comprising enhancing expression of an
CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
CC catalysing conversion of UDP-GlcNAc to mannose (Man)Nac, sialic acid
CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)
CC synthetase or CMP-SA transporter, where the expression of each enzyme is
CC enhanced to above endogenous levels. The novel method is useful for
CC manipulating glycoprotein production in an insect cell. Further methods
CC of the invention are useful for producing sialylated glycoprotein. The
CC sialylated glycoprotein produced by the above mentioned methods are
CC useful as pharmaceutical compositions, vaccines, diagnostics and
CC therapeutics. This polynucleotide sequence represents the DNA encoding a
CC human aldolase protein of the invention

XX SQ Sequence 1429 BP; 400 A; 284 C; 324 G; 421 T; 0 U; 0 Other;

Query Match 100.0%; Score 1429; DB 7; Length 1429;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCCAAGAGAAACTTTCAGGGTCTTGTGGCTGCAACCATCAGCCCAATGACT 60

Db 1 ATGGCTTCCCAAGAGAAACTTTCAGGGTCTTGTGGCTGCAACCATCAGCCCAATGACT 60

QY 61 GAGATGGAGAAATCAACTTTTCAGTAATTTGGTCAAGTATGTTGGATTTCTTGTGAAGAA 120

Db 61 GAGATGGAGAAATCAACTTTTCAGTAATTTGGTCAAGTATGTTGGATTTCTTGTGAAGAA 120

QY 121 CAGGAGTGAAGAACATTTTGTGAATGGCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180

Db 121 CAGGAGTGAAGAACATTTTGTGAATGGCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180

QY 181 TCAGAGCTGCGCCAGAGTTGCAGAGAGAGTGGTGCACAAAGAGAGAGAGAGAGAGAGAGAG 240

Db 181 TCAGAGCTGCGCCAGAGTTGCAGAGAGAGTGGTGCACAAAGAGAGAGAGAGAGAGAGAGAG 240

QY 241 GTGATTAATTCAGTAGAGACAGTGTGAGAGAGTGAAGAGTCAAGAGTGTGGCCCAACATGCA 300

Db 241 GTGATTAATTCAGTAGAGACAGTGTGAGAGTGAAGAGTCAAGAGTGTGGCCCAACATGCA 300

QY 301 GCAGAAATAGAGCTGATGGCATCGCTGTCATTTGCAAGAGAGAGAGAGAGAGAGAGAGAG 360

Db 301 GCAGAAATAGAGCTGATGGCATCGCTGTCATTTGCAAGAGAGAGAGAGAGAGAGAGAGAG 360

QY 361 AAAGATATCTGTATTAATTTCTTAAAGAGAGTGGTGTCTGCTGCTGCTGCTGCTGCTGCTG 420

Db 361 AAAGATATCTGTATTAATTTCTTAAAGAGAGTGGTGTCTGCTGCTGCTGCTGCTGCTGCTG 420

QY 421 TATTACTATCAGATTTCTGCTTGCAGAGAGTGAAGATTCGCTGCTGAGGAGTGTGGAT 480

Db 421 TATTACTATCAGATTTCTGCTTGCAGAGAGTGAAGATTCGCTGCTGAGGAGTGTGGAT 480

QY 481 GGGATTTCTGATAAGATCCCACTTCCAGAGGCTGAATTCAGTGTATCAGATCTCTTA 540

Db 481 GGGATTTCTGATAAGATCCCACTTCCAGAGGCTGAATTCAGTGTATCAGATCTCTTA 540

QY 541 GACTTCGGGCAATGTGTTGATCAGAAATCGCCAGCAACAGTTTCTTTTGGGGTG 600

Db 541 GACTTCGGGCAATGTGTTGATCAGAAATCGCCAGCAACAGTTTCTTTTGGGGTG 600

QY 601 GATGAGCAACTGTGTAGTGTCTGTGATGATGAGAGCAACTGAGCAGTGTGGTGTGTA 660

Db 601 GATGAGCAACTGTGTAGTGTCTGTGATGATGAGAGCAACTGAGCAGTGTGGTGTGTA 660

QY 661 TCCAGAGATTTATCAACTTTGTGCAAACTAGTGTGAGGTGTGAGGTGTGAGGTGTGAGGTGT 720

Db 661 TCCAGAGATTTATCAACTTTGTGCAAACTAGTGTGAGGTGTGAGGTGTGAGGTGTGAGGTGT 720

XX	SQ	Sequence 2536 BP; 744 A; 506 C; 500 G; 736 T; 0 U; 0 Other;	
		Query Match 93.2%; Score 1332.4; DB 3; Length 2536;	
		Best Local Similarity 94.3%; Pred. Kc. 0;	
		Matches 1428; Conservative 0; Mismatches 1; Indels 85; Gaps 1;	
QY	1	ATGGGCTTCCCAAGAGAAACTCAGGGTCTGTGGCTGCAACCAATCACGCCAATGACT 60	Db 1049
Db	89	ATGGGCTTCCCAAGAGAAACTCAGGGTCTGTGGCTGCAACCAATCACGCCAATGACT 148	QY 936
QY	61	GAGAAATGAGAAATCAACTTTTCAGTAATTCGTAGTATGTGGATATCTTSGAAAGAA 120	Db 1109
Db	149	GAGAAATGAGAAATCAACTTTTCAGTAATTCGTAGTATGTGGATATCTTSGAAAGAA 208	QY 996
QY	121	CAGGAGTGAAGACATTTTGTGATGCAACAGAGAGAGGCTGTCCCTGAGCGTC 180	Db 1169
Db	209	CAGGAGTGAAGACATTTTGTGATGCAACAGAGAGAGGCTGTCCCTGAGCGTC 268	QY 1169
QY	181	TCAGAGCGTCCGACAGGTTGCAGAGGAGTGGGTGCAAAAAGGGAAGCAAGCTGGATCAG 240	QY 996
Db	269	TCAGAGCGTCCGACAGGTTGCAGAGGAGTGGGTGCAAAAAGGGAAGCAAGCTGGATCAG 328	Db 1228
QY	241	GTGATTAATTCAGTAGGAGCACTGAGCTTGAAGAGTCA CAGAACTGCCCAACATGCA 300	QY 1055
Db	329	GTGATTAATTCAGTAGGAGCACTGAGCTTGAAGAGTCA CAGAACTGCCCAACATGCA 388	Db 1228
QY	301	GCAGAAATAGGAGCTGATGGCATCGCTGTCNTTGCACCGTCTTCTCCTCAAGCCNTGGACC 360	QY 1055
Db	389	GCAGAAATAGGAGCTGATGGCATCGCTGTCNTTGCACCGTCTTCTCCTCAAGCCNTGGACC 448	Db 1228
QY	361	AAAGATATCCTGATTAATTTCTTAAAGGAGTGGCTGCTGCCGCCCTGCTCCCTGCAATTT 420	QY 1115
Db	449	AAAGATATCCTGATTAATTTCTTAAAGGAGTGGCTGCTGCCGCCCTGCTCCCTGCAATTT 508	Db 1288
QY	421	TATTAATTAATTCCTCCCTGACAGGGTAAGATTCGTCTGAGAGTGTGTGGAT 480	QY 1175
Db	509	TATTAATTAATTCCTCCCTGACAGGGTAAGATTCGTCTGAGAGTGTGTGGAT 568	Db 1348
QY	481	GGGATTTCTGGATAGATCCCACTCCCAAGGCTGAATTCAGTACATACAGATCTCTTA 540	QY 1235
Db	569	GGGATTTCTGGATAGATCCCACTCCCAAGGCTGAATTCAGTACATACAGATCTCTTA 628	Db 1408
QY	541	GACTTCGGGCAATGTCTGATCAGATTCGACCAACAGTTTCTTCTTTTGGGGTG 600	QY 1235
Db	629	GACTTCGGGCAATGTCTGATCAGATTCGACCAACAGTTTCTTCTTTTGGGGTG 688	Db 1355
QY	601	GATGAGCAATCTGTAGTGTCTGTGTGATGGAGCAACTGGAGCACTGGG 550	Db 1528
Db	689	GATGAGCAATCTGTAGTGTCTGTGTGATGGAGCAACTGGAGCACTGGG 748	QY 1415
QY	651	----- 650	Db 1529
Db	749	ACTACTGGGAAAAAAGACAAACCCAGATGTGGAGGCTTTTGNACAAAGACATCTCT 808	QY 1415
QY	651	-----CAGTTTGTATTCAGAGATTTATCAACTTGTGTCAAACTAGGT 695	Db 1588
Db	809	TTAGCCCTGAACATATCAGTTTGTATCCAGAGATTTATCAACTTGTGTCAAACTAGGT 868	
QY	696	TTTGAGGTGTACAGACCAAGCCATCATGCTGTGGTCTCTGGGATTCATATGGGCCCA 755	
Db	869	TTTGAGGTGTACAGACCAAGCCATCATGCTGTGGTCTCTGGGATTCATATGGGCCCA 928	
QY	756	CCCGGCTTCCACTGCAGAAAGCTCCAGGAGTTTACTGATGAGTCTGAAGCTAAACTG 815	
Db	929	CCCGGCTTCCACTGCAGAAAGCTCCAGGAGTTTACTGATGAGTCTGAAGCTAAACTG 988	
QY	816	AAGAGCTGGATTTCTTTCTTCACTGATTTAAAGGATGGAACCTTGAAGCTGGTAGC 875	
Db	989	AAGAGCTGGATTTCTTTCTTCACTGATTTAAAGGATGGAACCTTGAAGCTGGTAGC 1048	
QY	876	TAGTGGCTCTCTATCAATCAGGGTTTGACCTTGAGACATATCTACCTTAATAGTGC 935	

RESULT 7	
ABN86096	
ID	ABN86096 standard; cDNA; 2441 BP.
XX	
AC	ABN86096;
XX	
DT	02-OCT-2002 (first entry)
XX	
DE	Polyadenylate binding protein 33.11 encoding cDNA.
XX	
KW	Polyadenylate binding 33.11; embryo development deformity; tumour;
KW	protein metabolism disorder; gene; ss.
XX	
OS	Unidentified.
XX	
Key	Location/Qualifiers
FT	318..1223
FT	/tag= a
FT	/product= "polyadenylate binding protein 33.11"
XX	
PN	CN1340520-A.
XX	
PD	20-MAR-2002.
XX	
PF	31-AUG-2000; 2000CN-00119821.
XX	
PR	31-AUG-2000; 2000CN-00119821.
XX	
PA	(BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX	
PI	Mao Y, Xie Y;
XX	
DR	WPI; 2002-436414/47.
DR	P-PSDB; ABB97988.

CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIFO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX

SQ Sequence 1171 BP; 302 A; 253 C; 339 G; 277 T; 0 U; 0 Other;
 Query Match 46.8%; Score 668.2; DB 4; Length 1171;
 Best Local Similarity 89.6%; Pred. No. 6.4e-181;
 Matches 788; Conservative 0; Mismatches 3; Indels 88; Gaps 3;

QY 1 ATGGCTTCCCAAGAGAACTTCAGGCTCTTGCTGCTGCAACCATCAGCCCAATGACT 60
 DB 192 ATGGCTTCCCAAGAGAACTTCAGGCTCTTGCTGCTGCAACCATCAGCCCAATGACT 251
 QY 61 GAGATGAGAAATCAACTTTTCAGTAATGCTCAGTATGCGATTATCTTCTGGAAGAA 120
 DB 252 GAGATGAGAAATCAACTTTTCAGTAATGCTCAGTATGCGATTATCTTCTGGAAGAA 311
 QY 121 CAGGAGTGAAGAAATTTTGTGATGCGCAACAGAGAGAGGCTGTCTCTGAGCGTC 180
 DB 312 CAGGAGTGAAGAAATTTTGTGATGCGCAACAGAGAGAGGCTGTCTCTGAGCGTC 371
 QY 181 TCAGAGCTGCGCAAGTTCAGAGAGTGGGTGACAAAGGAGGAGCAAGCTGATCAG 240
 DB 372 TCAGAGCTGCGCAAGTTCAGAGAGTGGGTGACAAAGGAGGAGCAAGCTGATCAG 431
 QY 241 GTGATAATTCAGTAGGAGCACTGAGCTTGAAGGAGTTCAGAGAACTGGCCCAATGCA 300
 DB 432 GTGATAATTCAGTAGGAGCACTGAGCTTGAAGGAGTTCAGAGAACTGGCCCAATGCA 491
 QY 301 GCAGAAATAGAGCTGNTGGCATCCTGTCTATGTCACCGTTCCTCTCAAGCATGGACC 360
 DB 492 GCAGAAATAGAGCTGNTGGCATCCTGTCTATGTCACCGTTCCTCTCAAGCATGGACC 551
 QY 361 AAAGATATCTGATTAATTTCTTAAGGAAGTGGTCTGCGCCCTGCGCTGCCATTT 420
 DB 552 AAAGATATCTGATTAATTTCTTAAGGAAGTGGTCTGCGCCCTGCGCTGCCATTT 511
 QY 421 TATTACTATCATTCCTGCTTGAAGGGTAAAGATTCGTGCTGAGGAGTTGTGGAT 480
 DB 612 TATTACTATCATTCCTGCTTGAAGGGTAAAGATTCGTGCTGAGGAGTTGTGGAT 671
 QY 481 GGGATCTGGATTAAGATCCCACTTCCAGGGCTGAATTCAGTATACATCTCTTA 540
 DB 672 GGGATCTGGATTAAGATCCCACTTCCAGGGCTGAATTCAGTATACATCTCTTA 731
 QY 541 GACTTCGGGCAATGCTTGTGATCAGAACTGCCAGCAAGTTCCTTCTTTCGGGGT 600
 DB 732 GACTTCGGGCAATGCTTGTGATCAGAACTGCCAGCAAGTTCCTTCTTTCGGGGT 791
 QY 601 GATGAGCAACTGTTGAGTGTCTTGTGATGGAGCAACTGGAGCAAGTGGG----- 650
 DB 792 GATGAGCAACTGTTGAGTGTCTTGTGATGGAGCAACTGGAGCAAGTGGGAGTACCTAT 851
 QY 651 ----- 650
 DB 852 AACTACCTGGGAAAAAGAACCAACAGATTTGTGGAGGCTTTTGAACAAAGGACTCTC 911
 QY 651 -----CAGTTTGTATCCAGAGATTTATCACTTTGTGTCAAACTAGG 694
 DB 912 TTTAGCCCTGAACTATCAGTTTGTATCCAGAGATTTATCACTTTGTGTCAAACTAGG 971
 QY 695 TTTTGGAGTGTCAAGAACCAAGCAATCATGCTGTGCTCT-GGGATTCCAATGGGCC 753

DB 972 TTTTGGAGTGTCAAGAACCAAGCCATCATGACTCTGCTCTGGGATTCATCGGCC 1031
 QY 754 CACCCCGGCTTCCAC-TGCAGAAAGCCTCCAGGGAGTTT 791
 DB 1032 CACCCCGGCTTCCACTTGCAGAAAGCCTCCAGGGAGTT 1070
 RESULT 9
 ABA04275
 ID ABA04275 standard; cDNA; 1483 BP.
 XX AC ABA04275;
 XX DT 04-MAR-2002 (first entry);
 XX DE Human dihydrodipyrindine synthetase 31 encoding cDNA SEQ ID NO:1.
 XX KW Human; dihydrodipyrindine synthetase 31; malignant tumour; noschaemia;
 XX KW HIV infection; immunological disease; inflammation; ss.
 XX CS Homo sapiens.
 XX FH Key Location/Qualifiers
 XX FT CDS 182..1036
 XX FT /*tag= a
 XX FT /product= "Human dihydrodipyrindine synthetase 31"
 XX PN CN1307108-A.
 XX XX 08-AUG-2001.
 XX XX 28-JAN-2000; 2000CN-00111575.
 XX XX 28-JAN-2000; 2000CN-00111575.
 XX XX (BODA-) BODAO GENE TECH CO LTD SHANGHAI.
 XX PI Mao Y, Xie Y;
 XX DR WPI; 2002-026890/04.
 XX DR P-PSDB; ABB04624.
 XX PT New polypeptide for treating malignant tumor and HIV infection, comprises
 XX PT the polypeptide-dihydrodi pyridine synthetase 31 and polynucleotide for
 XX PT coding said polypeptide.
 XX PS Claim 6; Page 25-26 (Disclosure); 34pp; Chinese.
 XX CC The present sequence encodes human dihydrodipyrindine synthetase 31 (I).
 XX CC (I) can be used in the treatment of various diseases, such as malignant
 XX CC tumour, noschaemia, HIV infection, immunological diseases and
 XX CC inflammations. The present invention also describes the antagonist
 XX CC resisting (I) and its treatment effect
 XX SQ Sequence 1483 BP; 420 A; 310 C; 344 G; 409 T; 0 U; 0 Other;

Query Match 45.7%; Score 652.4; DB 6; Length 1483;
 Best Local Similarity 99.8%; Pred. No. 2.5e-176;
 Matches 653; Conservative 0; Mismatches 1; Indels 3; Gaps 0;
 QY 1 ATGGCTTCCCAAGAGAACTTCAGGCTCTTGCTGCTGCAACCATCAGCCCAATGACT 60
 DB 182 ATGGCTTCCCAAGAGAACTTCAGGCTCTTGCTGCTGCAACCATCAGCCCAATGACT 241
 QY 61 GAGATGAGAAATCAACTTTTGTGATGCGCAACAGAGAGAGGCTGTCTCTGAGCGTC 120
 DB 242 GAGATGAGAAATCAACTTTTGTGATGCGCAACAGAGAGAGGCTGTCTCTGAGCGTC 301
 QY 121 CAGGAGTGAAGAAATTTTGTGATGCGCAACAGAGAGAGGCTGTCTCTGAGCGTC 180
 DB 302 CAGGAGTGAAGAAATTTTGTGATGCGCAACAGAGAGAGGCTGTCTCTGAGCGTC 361
 QY 181 TCAGAGCTGCGCAAGTTCAGAGAGTGGGTGACAAAGGAGGAGCAAGCTGATCAG 240

Db 362 TCAGAGCGTCGCCAGGTTGACAGAGAGTGGTGCACAAAGGAGAGCAGAGCTGGATCAG 421
 QY 241 GTGATTAATTCACGTAGAGCACTGAGCTTGAAGAGGTCAACAGGAACCTGGCCCAACATGCA 300
 Db 422 GTGATTAATTCACGTAGAGCACTGAGCTTGAAGAGGTCAACAGGAACCTGGCCCAACATGCA 481
 QY 301 GCAGAAATAGAGCTGATGTCATCGCTGTCATTCACCGCTTCTTCTCAAGCCATGGACC 360
 Db 482 GCAGAAATAGAGCTGATGTCATCGCTGTCATTCACCGCTTCTTCTCAAGCCATGGACC 541
 QY 361 AAGAGATATCCTGATTAATTTCTTAAGAGAGTGGCTGCTGCGCCCTGCGCCCTGCCATTT 420
 Db 542 AAGAGATATCCTGATTAATTTCTTAAGAGAGTGGCTGCTGCGCCCTGCGCCCTGCCATTT 601
 QY 421 TATTACTATCAGATTCCTGCTTGACAGGGGTAAAGATTGCTGCTGAGAGTGTGGAT 480
 Db 602 TATTACTATCAGATTCCTGCTTGACAGGGGTAAAGATTGCTGCTGAGAGTGTGGAT 661
 QY 481 GCGATTCTGGATAAGATCCCACTTCCAAAGGCTGAAATTCAGTGATACAGATCTCTTA 540
 Db 562 GCGATTCTGGATAAGATCCCACTTCCAAAGGCTGAAATTCAGTGATACAGATCTCTTA 721
 QY 541 GACTTCGGCAATGTGTGATCAGAAATGCGCCCAACAGTTTGCCTTTTGGGGTG 600
 Db 722 GACTTCGGCAATGTGTGATCAGAAATGCGCCCAACAGTTTGCCTTTTGGGGTG 781
 QY 601 GATGCAACTGTGATGCTGCTGATGGAGCACTGGAGCAGTGGGCAGT 654
 Db 782 GATGCAACTGTGATGCTGCTGATGGAGCACTGGAGCAGTGGGCAGT 835

RESULT 10

AAH52191

ID AAH52191 standard; cDNA; 444 BP.

XX

AC AAH52191;

XX

DT 10-SEP-2001 (first entry)

XX

DE Human APP protein encoding cDNA sequence SEQ ID NO:197.

XX

KW Human; secreted protein; secretion; bacterial cell; fungal cell;

KW eukaryotic cell; fusion protein; maltose binding protein;

KW immunoglobulin constant region; polyhistidine tag; ss.

XX

OS Homo sapiens.

XX

FN WO200129221-A2.

XX

PD 26-APR-2001.

XX

PF 20-OCT-2000; 2000WO-US029052.

XX

PR 20-OCT-1999; 99US-0160712P.

XX

PA (ZYMO) ZYMOGENETICS INC.

XX

PI Conklin DC, Yee DF;

XX

DR WPI; 2001-300340/31.

XX

DR P-PSDB; AAG81340.

XX

PT Isolated polypeptide for directing secretion of proteins of interest from a host cell including, e.g. bacteria, includes contiguous amino acid residues of polypeptide with specified amino acids.

XX

PS Claim 9; Page 347-348; 617pp; English.

XX

CC AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242 to AAG81453. The secreted proteins can be used for directing the

CC secretion of proteins of interest from a host cell including bacteria,

CC fungal cells, and cultured higher eukaryotic cells. The present invention

CC also describes fusion proteins, where a secreted protein of the invention is operably linked via a peptide bond or peptide linker to a second protein selected from the group consisting of maltose binding protein, an immunoglobulin constant region, a polyhistidine tag and a peptide given in AAG81453

XX

SQ Sequence 444 BP; 126 A; 97 C; 117 G; 104 T; 0 U; 0 Other;

Query Match

Best Local Similarity 30.3%; Score 433; DB 5; Length 444;

Matches 444; Conservative 99.8%; Pred. No. 1.3e-113;

Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGCGCTTCCCAAGAGAAACATTCAGGGTCTTGGCTGCAACATCAAGCCCAATGACT 60

Db 1 ATGCGCTTCCCAAGAGAAACATTCAGGGTCTTGGCTGCAACATCAAGCCCAATGACT 60

QY 61 GAGAAATGGAGAAATCAACTTTTCAGTAAATGGTCAGTATGGGATTTCTTGTGAAGAA 120

Db 61 GAGAAATGGAGAAATCAACTTTTCAGTAAATGGTCAGTATGGGATTTCTTGTGAAGAA 120

QY 121 CAGGGAGTGAAGAAACATTTTGTGAATGGTCACACAGGAGAGCGCTGTCCCTGAGGTC 180

Db 121 CAGGGAGTGAAGAAACATTTTGTGAATGGTCACACAGGAGAGCGCTGTCCCTGAGGTC 180

QY 181 TCAGAGCGTCGCCAGGTTSCAGAGGAGTGGTGCACAAAGAGGAGGACAGCTGGATCAG 240

Db 181 TCAGAGCGTCGCCAGGTTSCAGAGGAGTGGTGCACAAAGAGGAGGACAGCTGGATCAG 240

QY 241 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGAGTCACAGGAATGGCCCAACATGCA 300

Db 241 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGAGTCACAGGAATGGCCCAACATGCA 300

QY 301 GCAGAAATAGAGCTGATGGCATCGCTGTCATTTGACCGTTTCTTCTCAAGCCATGGACC 360

Db 301 GCAGAAATAGAGCTGATGGCATCGCTGTCATTTGACCGTTTCTTCTCAAGCCATGGACC 360

QY 361 AAAGATATCCTGATTAATTTCTTAAGAGAGTGGCTGCTGCGCCCTGCGCCCTGCCATTT 420

Db 361 AAAGATATCCTGATTAATTTCTTAAGAGAGTGGCTGCTGCGCCCTGCGCCCTGCCATTT 420

QY 421 TATTACTATCAGATTCCTGCTTGA 445

Db 420 TATTACTATCAGATTCCTGCTTGA 444

RESULT 11

AAK63371

ID AAK63371 standard; cDNA; 753 BP.

XX

AC AAK63371;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:8431.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer; cytostatic; gene therapy; vaccine; metastasis; ss.

XX

OS Homo sapiens.

XX

EN WO200157182-A2.

XX

XX 09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US001354.

XX

PR 31-JAN-2000; 2000US-0179065P.

XX

PR 04-FEB-2000; 2000US-0180628P.

XX

PR 24-FEB-2000; 2000US-0184664P.

XX

PR 02-MAR-2000; 2000US-0186350P.

XX

PR 15-MAR-2000; 2000US-0189874P.

XX

PR 17-MAR-2000; 2000US-0190076P.

XX

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0285515P.
PR 07-JUN-2000; 2000US-0239467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225271P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229539P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232082P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234223P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0243207P.
PR 17-NOV-2000; 2000US-0243208P.
PR 17-NOV-2000; 2000US-0243209P.
PR 17-NOV-2000; 2000US-0243210P.
PR 17-NOV-2000; 2000US-0243211P.
PR 17-NOV-2000; 2000US-0243212P.
PR 17-NOV-2000; 2000US-0243213P.
PR 17-NOV-2000; 2000US-0243214P.
PR 17-NOV-2000; 2000US-0243215P.
PR 17-NOV-2000; 2000US-0243216P.
PR 17-NOV-2000; 2000US-0243217P.
PR 17-NOV-2000; 2000US-0243218P.
PR 17-NOV-2000; 2000US-0243244P.
PR 17-NOV-2000; 2000US-0243245P.
PR 17-NOV-2000; 2000US-0243246P.
PR 17-NOV-2000; 2000US-0243265P.
PR 17-NOV-2000; 2000US-0243297P.
PR 17-NOV-2000; 2000US-0243299P.
PR 17-NOV-2000; 2000US-0243300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WFI; 2001-483426/52.
XX P-PSDB; AAM90590.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides;
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Claim 1; SEQ ID NO 8431; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and


```
PR 09-JAN-2002; 2002US-0347313P.
XX (CORI-) CORIXA CORP.
PA Wang T, Wang S, Bangur CS, Gaiger A;
PI WPI; 2003-058387/05.
XX
DR New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and in
XX virology, immunology, microbiology, molecular biology and recombinant DNA
XX techniques.
PS Claim 1; SEQ ID NO 1700; 207pp; English.
XX
XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX ABP54446 to ABP54472 represent protein (II) sequences, from the present
XX invention. (I) and (II) have cytostatic activity and can be used in gene
XX therapy and vaccines. (I), (II), antibodies and compositions from the
XX present invention are useful for diagnosing, preventing and treating
XX cancer, which expresses CT or CP mRNA antigens. They are useful for
XX stimulating immune response. They can also be useful in virology,
XX immunology, microbiology, molecular biology and recombinant DNA
XX techniques. N.B. The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 628 BP; 191 A; 123 C; 143 G; 171 T; 0 U; 0 Other;

Query Match      8.7%; Score 124; DB 7; Length 628;
Best Local Similarity 100.0%; Pred. No. 7.9e-25;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 531 AGATCTCTTAGACTTCGGCAATGTTGATCAGAAATGCCAGCAACAGTTTCCT 590
Db 628 AGATCTCTTAGACTTCGGCAATGTTGATCAGAAATGCCAGCAACAGTTTCCT 569

Qy 591 TTTGGGGTGGATGAGCAACTGTTGAGTCTCTGATGGAGCACTGGAGCAGTGGG 650
Db 568 TTTGGGGTGGATGAGCAACTGTTGAGTCTCTGATGGAGCACTGGAGCAGTGGG 509

Qy 651 CAGT 654
Db 508 CAGT 505

RESULT 14
AAK75386
XX ID AAK75386 standard; DNA; 14504 BP.
XX AC AAK75386;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:30198.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 27-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226273P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233633P.
PR 14-SEP-2000; 2000US-0233634P.
PR 14-SEP-2000; 2000US-0233655P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237337P.
PR 02-OCT-2000; 2000US-0237338P.
PR 02-OCT-2000; 2000US-0237339P.
PR 13-OCT-2000; 2000US-0237340P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
```

PR 20-OCT-2000; 2000US-0240360P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 XX
 PS Disclosure; SEQ ID NO 30198; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and

CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to prevent the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK37694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 14504 BP; 4313 A; 3094 C; 3219 G; 4278 T; 0 U; 0 Other;
 Query Match 6.7%; Score 95.4; DB 4; Length 14904;
 Best Local Similarity 99.0%; Pred. No. 6.7e-16;
 Matches 96; Conservative 0; Mismatches 1; Indels 0; Gaps 3;
 QY 363 AGATATCCTGATTAATTTCTTAAAGGAAGTGGCTGCGCGCCCTGCCCTGCCATTTTA 422
 DB 13017 AGATATCCTGATTAATTTCTTAAAGGAAGTGGCTGCGCGCCCTGCCCTGCCATTTTA 13076
 QY 423 TTACTATCATTCCTGCTTGTGACAGGGTAAAGATT 459
 DB 13077 TTACTATCATTCCTGCTTGTGACAGGGTAAAGATT 13113
 RESULT 15
 ABA09731
 ID ABA09731 standard; DNA; 618 BP.
 XX ABA09731;
 XX
 DT 15-JAN-2002 (first entry)
 XX
 DE Human bone marrow expressed oligonucleotide SEQ ID NO: 240.
 XX
 KW Human; bone marrow; cytostatic; antirheumatic; antiarthritic; vulnary;
 KW antinflammatory; antibacterial; immunosuppressive; vasotropic; cancer;
 KW antiparkinsonian; neuroprotective; nootropic; haemostatic; osteopathic;
 KW antitumor; fungicide; antidiabetic; antisthmatic; antiallergic;
 KW immunostimulant; analgesic; cerebroprotective; antianaemic; infection;
 KW nervous system disorder; autoimmune disorder; inflammation; allergy; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200174836-A1.
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US010472.
 XX
 XX 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 PR 30-NOV-2000; 2000US-0250583P.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT, Ford JE, Boyle BJ;
 XX WPI; 2001-626375/72.
 DR
 XX
 PT New bone marrow-expressed nucleic acids and polypeptides, useful for
 PT diagnosis, treatment of inflammatory, autoimmune, neurological, cancer
 PT and increasing hematopoiesis, stem cell survival and bone growth and
 PT remodeling.
 XX
 XX Claim 1; Page 299; 380pp; English.
 PS
 CC The present invention relates to bone marrow expressed polynucleotides
 CC and proteins. These sequences can be used in the treatment of

CC inflammatory conditions (eg arthritis, Crohn's disease), cancer, central
 CC and peripheral nervous system diseases and neuropathies, such as
 CC Alzheimer's, Parkinson's and Huntington's diseases, spinal cord
 CC disorders, head trauma, cerebrovascular diseases, myeloid and lymphoid
 CC cell disorders, platelet disorders, stem cell disorders, bone
 CC degenerative disorders, autoimmune disorders, for example multiple
 CC sclerosis, diabetes and arthritis, viral and bacterial infections,
 CC allergies and blood coagulation disorders. The present sequence is a DNA
 CC of the invention
 XX

SQ Sequence 618 BP; 186 A; 143 C; 155 G; 134 T; 0 U; 0 Other;

Query Match 5.4%; Score 77.2; DB 5; Length 618;
 Best Local Similarity 96.3%; Pred. No. 2.1e-11;
 Matches 79; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	65	ATGGGAAATCAACTTTTCAGTAAATGGTCAGTATCTGGATTATCTGTGAAGAACAGG	124
DB	254	ACGGGAAATCAACTTTTCAGTAAATGGTCAGTATCTGGATTATCTGTGAAGAACAGG	313
QY	125	GAGTGAAGAACATTTTGTGA	146
DB	314	GAGTGAAGAACATTTTGTGA	335

Search completed: May 10, 2004, 08:07:50
 Job time : 629.115 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:33 ; Search time 4131.87 Seconds
(without alignments)
10327.779 Million cell updates/sec

Title: US-09-930-440B-1

Perfect score: 1429

Sequence: 1 atggcctcccaagaagaa.....ctttgaattcatttcgatg 1429

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estit:*
4: em_estim:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852.2	59.6	1015	9 AL525867	AL525867
2	698.8	48.9	1201	9 AL577521	AL577521
3	671.8	47.0	1201	9 AL554919	AL554919
4	670.8	46.9	859	12 B1760786	B1760786

5	654	45.8	951	9	AL545960
6	643	45.0	1036	9	AL541544
7	637.2	44.6	1747	11	BC042003
8	608.2	42.6	709	14	CA423436
9	592.8	41.5	1403	11	AK088859
10	588.2	41.2	1403	11	AK002734
11	587.4	41.1	654	14	CK001323
12	537.2	37.6	753	9	AV728536
13	531.4	37.2	643	13	BU658094
14	531.4	37.2	893	14	CB999084
15	528	36.9	723	14	CB985894
16	522.4	36.6	625	9	AL597631
17	508	35.5	508	9	AI916625
18	507	35.5	508	9	AI521193
19	503.6	35.2	790	14	CK020463
20	503.6	35.2	1006	13	BY703240
21	500	35.0	635	12	BG426736
22	499.6	35.0	731	12	BI144011
23	498	34.8	975	13	BY747131
24	497.8	34.8	704	14	CD366774
25	496.2	34.7	693	12	BM941869
26	493.2	34.5	834	12	BI647729
27	488.6	34.2	691	14	CB057833
28	484	33.9	496	10	BF740024
29	478	33.4	511	14	CA396517
30	464.2	32.5	604	14	CD704049
31	463.4	32.4	465	9	AI635718
32	463.4	32.4	859	12	BI832112
33	459	32.1	890	10	BF691187
34	447.4	31.3	758	14	CB571304
35	446.4	31.2	774	14	CB231745
36	443	31.0	444	14	W78156
37	426.4	29.8	2412	11	SC034966
38	425.6	29.8	730	14	CB600330
39	420.8	29.4	876	13	BU199204
40	417.4	29.2	422	10	BF876452
41	412	28.8	646	13	BY739779
42	398	27.9	620	14	CD693284
43	392.6	27.5	669	12	BG297148
44	390.8	27.3	503	14	T87364
45	390.6	27.3	398	12	BM989844

ALIGNMENTS

RESULT 1

AL525867 1015 bp mRNA linear EST 23-MAY-2003
Locus: AL525867 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
Definition: CDNA clone CSODC013YK06 5-PRIME, mRNA sequence.
Accession: AL525867
Version: AL525867.2 GI:31063731
Keywords: EST.
Source: Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 1015)
Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Full-length cDNA libraries and normalization
Unpublished (2001)

AUTHORS

TITLE

JOURNAL

COMMENT

On Feb 13, 2001 this sequence version replaced gi:12789360.

Contact: Genoscope

Genoscope - Centre National de Sequenage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 291.f For

more information about this cluster, see

<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODC013BF030P1&cluster=291.f>

Contact : Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1693
 Faraçay Avenue Genoscope sequence ID : CS0DC013BF03QPL.

FEATURES

```

1. 1015
location/qualifiers
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="UCSDC013YK06"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NcoI-oligo (GT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

ORIGIN

Query Match 53.6%; Score 852.2; DB 9; Length 1015;
Best Local Similarity 59.8%; Pred. No. 7.1e-232;
Matches 851; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1	QY	ATGCCCTTCCCAAGAGAAACTTCAGGGTCTTGTGGCTGCAACCATACAGCCCAATGACT	60
161	Db	ATGCCCTTCCCAAGAGAAACTTCAGGGTCTTGTGGCTGCAACCATACAGCCCAATGACT	220
61	QY	GNGAATCGAGAAATCAACTTTTCAGTAACTGGTCAGTATGTGGATATCTTCTGAAAGAA	120
221	Db	GNGAATCGAGAAATCAACTTTTCAGTAACTGGTCAGTATGTGGATATCTTCTGAAAGAA	280
121	QY	CAGGGAGTGAAGAACNTTTTGTGAATGSCACACAGGAGAGGGCTGTCCCTGAGCGTC	180
281	Db	CAGGGAGTGAAGAACNTTTTGTGAATGSCACACAGGAGAGGGCTGTCCCTGAGCGTC	340
181	QY	TCAGAGCGTCGCCAGGTTGCAGAGAGTGGGTGACAAAGGGAAGACAGCTGATCAG	240
341	Db	TCAGAGCGTCGCCAGGTTGCAGAGAGTGGGTGACAAAGGGAAGACAGCTGATCAG	400
241	QY	GTGATAATTACGTPAGAGCACTGAGCTTTGAAGGAGTCACAGGAATGCGCCCAACATGCA	300
401	Db	GTGATAATTACGTPAGAGCACTGAGCTTTGAAGGAGTCACAGGAATGCGCCCAACATGCA	460
301	QY	GCAGAAATAGAGCTGATGGCATCGCTCTCATTTGCACCGTTCTTCTCAAGCCATGGACC	360
461	Db	GCAGAAATAGAGCTGATGGCATCGCTCTCATTTGCACCGTTCTTCTCAAGCCATGGACC	520
361	QY	AAAGATATCTGATTAATTTCTTAAGGAAGTGGCTGCTCGGCCCTCGCCCTGCGATT	420
521	Db	AAAGATATCTGATTAATTTCTTAAGGAAGTGGCTGCTCGGCCCTCGCCCTGCGATT	580
421	QY	TAATTACTATCACATTTCTCGCTTGACAGGGCTAAAGATTCTGTCGAGAGTGTGTGGAT	480
581	Db	TAATTACTATCACATTTCTCGCTTGACAGGGCTAAAGATTCTGTCGAGAGTGTGTGGAT	640
481	QY	GGGATTTCTGATAGATCCCACTTCCAGGGCTGAAATTCAGTGAATACAGATCTCTTA	540
641	Db	GGGATTTCTGATAGATCCCACTTCCAGGGCTGAAATTCAGTGAATACAGATCTCTTA	700
541	QY	GACTTCGGGCAATGCTGATCAAGATCGCCAGCAACNGTTTGCTTTCTTTTGGGGTG	600
701	Db	GACTTCGGGCAATGCTGATCAAGATCGCCAGCAACNGTTTGCTTTCTTTTGGGGTG	760
601	QY	GATGAGCAACTGTTGAGTGTCTTGGTATGGGAGCACTGGAGGAGTGGGAGTTTGTGA	660
761	Db	GATGAGCAACTGTTGAGTGTCTTGGTATGGGAGCACTGGAGGAGTGGGAGTTTGTGA	820
661	QY	TCAGAGATTATCAACTTTGTGTGCAAACTAGGTTTTTGGAGTGTCAAGACCAAGCCA	720
821	Db	TCAGAGATTATCAACTTTGTGTGCAAACTAGGTTTTTGGAGTGTCAAGACCAAGCCA	880
721	QY	TCATGACTCTGGTCTCTGGGATTTCCAAATGGGCCCAACCCCGGCTTCCATCTGCAGAAAGCT	780
881	Db	TCATGACTCTGGTCTCTGGGATTTCCAAATGGGCCCAACCCCGGCTTCCATCTGCAGAAAGCT	940
781	QY	CCAGGGAGTTTACTGATAGTGTGAAGCTAACTGAAGAGCTGTGATTTCTTTCTTTCA	840

```
Db 778 RSCITTTGAACAAGAAAGTCTCTCTTTAGCCCTGAACATACAGTTTGTATCCAGAGT 719
Qy 671 TATCAACTTTTGTGTAAGTCTTTGGAGTGTCAAGACCAAGCCATCATGCTCT 730
Db 718 TATCAACTTTTGTGTAAGTCTTTGGAGTGTCAAGACCAAGCCATCATGCTCT 659
Qy 731 GGTCTCTGGGATTCAGTGGGCGCCACCGG-GCTTCCGATCGAGAAAGCCCTCCAGGAGT 789
Db 658 GGTCTCTGGGATTCAGTGGGCGCCACCGGCGCTTCCGATCGAGAAAGCCCTCCAGGAGT 599
Qy 790 TTACTGATAGTCTGAAGCTAAAGTGAAGAGAGCTGATCTCTCTTTTCACTGATTTAA 849
Db 598 TNACTGATAGTCTGAAGCTAAAGTGAAGAGAGCTGATCTCTCTTTTCACTGATTTAA 539
Qy 850 AGGATGGAACTTTGTAAGGCGGAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 909
Db 538 AGGATGGAACTTTGTAAGGCGGAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 479
Qy 910 GAGACATAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 969
Db 478 GAAACATAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 419
Qy 970 TAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1029
Db 418 TAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 359
Qy 1030 AAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1089
Db 358 AAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 299
Qy 1090 TCAGAGATTTTGTGAGAGAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1149
Db 298 TCAGAGATTTTGTGAGAGAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 239
Qy 1150 ATTGTAATTTGTAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1209
Db 238 ATTGTAATTTGTAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 179
Qy 1210 TAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1269
Db 178 TAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 119
Qy 1270 TAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1328
Db 118 TAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 59
Qy 1329 CAAACTGTAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1383
Db 58 CAAACTGTAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4
```

RESULT 3

```
AL554919
LOCUS AL554919 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION Clone CS0D1087YB23 5-PRIME, mRNA sequence.
ACCESSION AL554919
VERSION AL554919.2 GI:31276729
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
L.I.W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:12896160.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
```

Invitrogen. This sequence belongs to sequence cluster 291.f For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1087YB23&cluster=291.f>. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Paraday Avenue Genoscope sequence ID : CS0D1087YB23P1. Location/Qualifiers

```
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1087YB23"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
```

FEATURES

source

ORIGIN

```
Query Match 47.0%; Score 671.8; DB 9; Length 1201;
Best Local Similarity 88.9%; Pred. No. 2.8e-180;
Matches 792; Conservative 4; Mismatches 7; Indels 88; Gaps 3;

Qy 465 TGAGGAGTTGTTGGATGGGATTCGGATTAAGATCCCGACCTTCGAGGCTGAATTCAG 524
Db 69 TGAGGAGTTGTTGGATGGGATTCGGATTAAGATCCCGACCTTCGAGGCTGAATTCAG 128
Qy 525 TGATACAGATCTCTTAGACTTCGGGCAATGCTTGATCAGATCGCACCAACAGTTTC 594
Db 129 TGATACAGATCTCTTAGACTTCGGGCAATGCTTGATCAGATCGCACCAACAGTTTC 188
Qy 585 TTTCTTTTGGGTTGGATGAGCAACTGTTGAGTCTCTGGTGGTGGGAGCAACTGGAGC 644
Db 189 TTTCTTTTGGGTTGGATGAGCAACTGTTGAGTCTCTGGTGGTGGGAGCAACTGGAGC 248
Qy 645 AGTGGG----- 650
Db 249 AGTGGGAGTACCTATAACTACCTGGGAAAAAAGACAAACCCAGATGTTGGAGCTTTGA 308
Qy 651-----CAGTTTGTATCCAGAGATTTATCAACT 679
Db 309 ACAAAGAGACTCTCTTTAGCCCTGAATCAGTTTGTATCCAGAGATTTATCAACT 368
Qy 680 TGTGTCAAACTAGGTTTGGAGTGTACAGACCAAGCCATCATGCTCTGCTCTGG 739
Db 369 TGTGTCAAACTAGGTTTGGAGTGTACAGACCAAGCCATCATGCTCTGCTCTGG 428
Qy 740 GATTCCAAATGGGCGCCACCGGCTTCCACTGAGAAAGCCCTCCAGGAGTTTACTGATAG 799
Db 429 GATTCCAAATGGGCGCCACCGGCTTCCACTGAGAAAGCCCTCCAGGAGTTTACTGATAG 488
Qy 800 TGCTGAAGCTAACTGAAGAGCTGGAATTCCTTTCTTTCTCTCTCTCTCTCTCTCTCTCT 859
Db 489 TGCTGAAGCTAACTGAAGAGCCCTGGAATTCCTTTCTTTCTCTCTCTCTCTCTCTCTCT 548
Qy 860 CTTGGAAGCTGGTACTAGTGCCTCTCTATCAAAATCAGGGTTTGACCTTGAGACATAT 919
Db 549 CTTGGAAGCTGGTACTAGTGCCTCTCTATCAAAATCAGGGTTTGACCTTGAGACATAT 608
Qy 920 CTACCTTAAATAGTGCATTTTCTCTGAGGAATTTAGATGAACTTGAATTAACCTCTCC 979
Db 609 CTACCTTAAATAGTGCATTTTCTCTGAGGAATTTAGATGAACTTGAATTAACCTCTCC 668
Qy 980 TAGCAATGAATCTCAACAATAAGCAATGAGTACTCTTTGTGAGCCCTTAAAGTCTTAA 1039
Db 669 TAGCAATGAATCTCAACAATAAGCAATGAGTACTCTTTGTGAGCHTTAAAGTCTTAA 728
Qy 1040 TTTTGTGAAGGGGCAAAAACCTCTAGGAGTCAAACTCTCAGTCACTTCATTTACAGATTT 1099
Db 729 TTTTGTGAAGGGGCAAAAACCTCTAGGAGTCAAACTCTCAGTCACTTCATTTACAGATTT 788
```

```

QY 1100 TTTTCGGAGAAATTTCTCTTTTATAGGATGAAATCAAGAGAAATGTAAATG 1159
Db 789 TTTTCGGAG-AATTTCTCTTTATAGGATGAAATCAAGAGAAATGTAAATG 847
QY 1160 ATTAATCCATCTGCTTTAGGAGCTCTCATATCTCGCTCTCTGGTTCCTAACTT 1219
DE 848 ATAACTCATCTGCTTTAGGAGCTCTCATATCTCGCTCTCTGGTTCCTAACTT 907
QY 1220 TTAAGTCTCTAATTTAAACACTATATATGCTCTCATTTTAAATAAT 1270
Db 908 TTAAGTCTCTAATTT--AACACTATATATGCTCTCATTTTAAATAAT 956

RESULT 4
LOCUS B1760786 859 bp mRNA linear EST 25-SEP-2001
DEFINITION 603043807F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5184192 5',
mRNA sequence.
ACCESSION B1760786
VERSION B1760786.1 GI:15752364
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11459 row: n column: 01
High quality sequence stop: 858.
FEATURES
Location/Qualifiers
1..853
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5184192"
/lab_host="DH10B"
/clone_lib="NIH MGC 116"
/note="Organ: pCooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."
ORIGIN
Query Match 46.3%; Score 670.8; DB 12; Length 859;
Best Local Similarity 97.5%; Pred. No. 4.6e-180;
Matches 745; Conservative 0; Mismatches 12; Indels 7; Gaps 6;
QY 1 ATGGCTTCCCAAGAGAACTTCAGGCTTTGTGCTGCAACCATCAGCCCAATGACT 60
Db 99 ATGGCTTCCCAAGAGAACTTCAGGCTTTGTGCTGCAACCATCAGCCCAATGACT 158
QY 61 GAGAAATGGAGAAATCACTTTTCAGTAATTTGGTCAGTATGTGATTAATCTTTGGAAGAA 120
Db 159 GAGAAATGGAGAAATCACTTTTCAGTAATTTGGTCAGTATGTGATTAATCTTTGGAAGAA 218

```

```

QY 121 CAGGGAGTGAAGAACATTTTGTGATGGCAACACAGAGAGAGCCCTGTCCCTGAGGCTC 180
Db 219 CAGGGAGTGAAGAACATTTTGTGATGGCAACACAGAGAGAGCCCTGTCCCTGAGGCTC 277
QY 181 TCAGAGCGTCGCCAGCTTCAGAGGAGTGGTGACAAAAGGGAAGACAAAGCTGGAATCAG 240
Db 278 TCAGAGCGTCGCCAGCTTCAGAGGAGTGGTGACAAAAGGGAAGACAAAGCTGGAATCAG 337
QY 241 GTGATAATTCAGTAGGACATGAGCTTTGAAGGAGTCACAGGAATGGGCCCAACATGCA 300
Db 338 GTGATAATTCAGTAGGACATGAGCTTTGAAGGAGTCACAGGAATGGGCCCAACATGCA 397
QY 301 GCAGAAATAGGAGCTGATGGCATGGCTGTCAATTCACCGTCTCTCTCAAGCCATGGACC 360
Db 398 GCAGAAATAGGAGCTGATGGCATGGCTGTCAATTCACCGTCTCTCTCAAGCCATGGACC 457
QY 361 AAAGATATCTCTGATTAATTTCTTAAAGAGAGTGGCTGTGCTGCCGCCCTGCTGCCCATTT 420
Db 458 AAAGATATCTCTGATTAATTTCTTAAAGAGAGTGGCTGTGCTGCCGCCCTGCTGCCCATTT 517
QY 421 TATTACTATCATTTCTGCTTGAAGAGTAAAGATTCGCTGCTGAGGAGTTGTTGGAT 480
Db 518 TATTACTATCATTTCTGCTTGAAGAGTAAAGATTCGCTGCTGAGGAGTTGTTGGAT 576
QY 481 GGGATCTCGATAGATCCCACTTCCAGGGGCTGAATTCAGTGTACAGATCTCTTA 540
Db 577 GGGATCTCGATAGATCCCACTTCCAGGGGCTGAATTCAGTGTACAGATCTCTTA 636
QY 541 GACTTCGGGCAATGTGTTGATCAGAAATCCAGCAACAGTTTCTTTCTTTTTCCTTTT 599
Db 637 GACTTCGGGCAATGTGTTGATCAGAAATCCAGCAACAGTTTCTTTCTTTTTCCTTTT 696
QY 600 GATGAGCAACTGTTCAGTGTCTGTGATGGAGCAACT-GGAGAGTGGGAGTTTG 658
Db 697 GGATGAGCAACTGTTCAGTGTCTGTGATGGAGCAACTGGGAGCAGTGGGAGTTTG 756
QY 659 TATCCAGAGATTTATCAACTTTTGTGCAAACTAGGTTTGGAGTGTACAGACCAAGC 718
Db 757 TATCCAGAGATTTATCAACTTTTGTGCAAACTAGGTTTGGAGTGTACAGACCAAGC 815
QY 719 CATCATGACTCTGG--TCTTGGGATTCGAATGGGCCCAACCCG 760
Db 816 CATCATGACTCTGGGTCCTCGGGATTCGAATGGGCCCAACCCG 859

RESULT 5
LOCUS AL546963
DEFINITION AL546960 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
clone CS0D1032XJ10 5-PRIME, mRNA sequence.
ACCESSION AL546960
VERSION AL546960.2 GI:31268793
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 951)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:12880585.
Contact: Genoscope - Centre National de Sequencage
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 291.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1032DE05QF1&cluster=291.f. Contact :
Feng Liang Email : fliang@life.techn.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

```

```

FEATURES
  source
    Faraday Avenue Genoscope sequence ID : CS0DI032DE05QPL.
    Location/Qualifiers
      1..951
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CS0DI032YU10"
        /tissue type="PLACENTA COT 25-NORMALIZED"
        /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
        /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
  Query Match 45.8%; Score 654; DB 9; Length 951;
  Best Local Similarity 100.0%; Pred. No. 3.1e-175;
  Matches 654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCAAGAGAACTTCAGGCTCTTGGCTGCAACCATCAAGCCAAATGACT 60
Db |||||||
QY 257 ATGGCCCTCCCAAGAGAACTTCAGGCTCTTGGCTGCAACCATCAAGCCAAATGACT 316
Db |||||||

QY 61 GAGAAATGGAGAAATCACTTTTCAGTAATTCGTCAATATGCTATGCTATCTTGTGAAGAA 120
Db |||||||
QY 317 GAGAAATGGAGAAATCACTTTTCAGTAATTCGTCAATATGCTATGCTATCTTGTGAAGAA 376
Db |||||||

QY 121 CAGGAGTGAAGAACATTTTGTGAATGGCAACAGGAGAGAGGCTTCCTCGAGCGTC 180
Db |||||||
QY 377 CAGGAGTGAAGAACATTTTGTGAATGGCAACAGGAGAGAGGCTTCCTCGAGCGTC 436
Db |||||||

QY 181 TCAGAGCGTCGCCAGGTGCAAGAGAGTGGTGTGCAAAAGGAGAGGACAGCTGATCAG 240
Db |||||||
QY 437 TCAGAGCGTCGCCAGGTGCAAGAGAGTGGTGTGCAAAAGGAGAGGACAGCTGATCAG 496
Db |||||||

QY 241 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGTCAACAGGAAGTGGCCCAACATGCA 300
Db |||||||
QY 497 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGTCAACAGGAAGTGGCCCAACATGCA 556
Db |||||||

QY 301 GCAGAAATAGAGCTGATGCGTGTCTATTCGACCGTTCCTCAAGCCATGGACC 360
Db |||||||
QY 557 GCAGAAATAGAGCTGATGCGTGTCTATTCGACCGTTCCTCAAGCCATGGACC 616
Db |||||||

QY 361 AAGAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCCGCCCTGCTGCCATTT 420
Db |||||||
QY 617 AAGAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCCGCCCTGCTGCCATTT 676
Db |||||||

QY 421 TATTACTATCATTCTCGCTTGCAGAGGTAAAGATTTCGTGAGGAGTGTGAT 480
Db |||||||
QY 677 TATTACTATCATTCTCGCTTGCAGAGGTAAAGATTTCGTGAGGAGTGTGAT 736
Db |||||||

QY 481 GGGATTCTGATTAAGATCCCACTTCCAGAGGCTGAAATTCAGTGATACAGATCTCTTA 540
Db |||||||
QY 737 GGGATTCTGATTAAGATCCCACTTCCAGAGGCTGAAATTCAGTGATACAGATCTCTTA 796
Db |||||||

QY 541 GACTTCGGGCAATGTGTGATCAGATTCGACCAACAGTTTCTTCTTTTGGGGTG 600
Db |||||||
QY 797 GACTTCGGGCAATGTGTGATCAGATTCGACCAACAGTTTCTTCTTTTGGGGTG 856
Db |||||||

QY 601 GATGAGCAATGTGTGATGCTCTGGTGTGAGGAGCAACTGGAGAGTGGGCACT 654
Db |||||||
QY 857 GATGAGCAATGTGTGATGCTCTGGTGTGAGGAGCAACTGGAGAGTGGGCACT 910
Db |||||||

RESULT 6
AL541544 1036 bp mRNA linear EST 12-MAY-2003
LOCUS AL541544 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE003YCI9
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION AL541544
VERSION AL541544.2 GI:30545825
KEYWORDS EST.
SOURCE Homo sapiens (human)

```

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1036)

AUTHORS

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

On Feb 15, 2001 this sequence version replaced gi:12872718.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 291.f For

more information about this cluster, see

http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DE003AB10P1&cluster=291.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DE003AB10QPL.

FEATURES

Location/Qualifiers

1..1036

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DE003YCI9"

/tissue type="PLACENTA"

/clone_lib="Homo sapiens PLACENTA"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo (dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

ORIGIN

Query Match 45.0%; Score 643; DB 9; Length 1036;

Best Local Similarity 99.8%; Pred. No. 4.4e-172;

Matches 654; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGGCCCTCCCAAGAGAACTTCAGGCTCTTGGCTGCAACCATCAAGCCAAATGACT 60

Db 234 ATGGCCCTCCCAAGAGAACTTCAGGCTCTTGGCTGCAACCATCAAGCCAAATGACT 293

QY 61 GAGAAATGGAGAAATCACTTTTCAGTAATTCGTCAATATGCTATGCTATCTTGTGAAGAA 120

Db 294 GAGAAATGGAGAAATCACTTTTCAGTAATTCGTCAATATGCTATGCTATCTTGTGAAGAA 353

QY 121 CAGGAGTGAAGAACATTTTGTGAATGGCAACAGGAGAGGCTTCCTCGAGCGTC 180

Db 354 CAGGAGTGAAGAACATTTTGTGAATGGCAACAGGAGAGGCTTCCTCGAGCGTC 413

QY 181 TCAGAGCGTCGCCAGGTGCAAGAGTGGTGTGCAAAAGGAGAGGACAGCTGATCAG 240

Db 414 TCAGAGCGTCGCCAGGTGCAAGAGTGGTGTGCAAAAGGAGAGGACAGCTGATCAG 473

QY 241 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGTCAACAGGAAGTGGCCCAACATGCA 300

Db 474 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGTCAACAGGAAGTGGCCCAACATGCA 533

QY 301 GCAGAAATAGAGCTGATGCGTGTCTATTCGACCGTTCCTCAAGCCATGGACC 360

Db 534 GCAGAAATAGAGCTGATGCGTGTCTATTCGACCGTTCCTCAAGCCATGGACC 593

QY 361 AAGAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCCGCCCTGCTGCCATTT 420

Db 594 AAGAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCCGCCCTGCTGCCATTT 653

QY 421 TATTACTATCATTCTCGCTTGCAGAGGTAAAGATTTCGTGAGGAGTGTGAT 479

Db 654 TATTACTATCATTCTCGCTTGCAGAGGTAAAGATTTCGTGAGGAGTGTGAT 713

QY 480 TGGGATTCGGATTAAGATCCCACTTCCAGGGCTGAAATTCAGTGATACAGATCTCTT 539


```

Db      714 TGGATTCTGGATAGATCCCACTTCCAAAGGGCTGAAATTCAGTGATACAGATCTCTT 773
Qy      540 AGACTTCGGGCAATGTTGATCAGAAATCGGCAGCAACAGTTTCTTTTCTTTTGGGGT 599
Db      774 AGACTTCGGGCAATGTTGATCAGAAATCGGCAGCAACAGTTTCTTTTCTTTTGGGGT 833
Qy      600 GGATGAGCACTGTGTAGTCTCTGTGATGGAGCACTGGAGCAGTGGCGAGT 654
Db      834 GGAATGAGCACTGTGTAGTCTCTGTGATGGAGCACTGGAGCAGTGGCGAGT 888

RESULT 7
BC042003      1747 bp      mRNA      linear      HTC 02-JAN-2003
DEFINITION    Homo sapiens, clone IMAGE:5311213, mRNA.
ACCESSION     BC042003
VERSION       BC042003.1 GI:27469544
KEYWORDS      HTC.
SOURCE        Homo sapiens (human).
ORGANISM      Homo sapiens
REFERENCE     1. (bases 1 to 1747)
AUTHORS       Strausberg, J.
JOURNAL       Submitted (23-DEC-2002) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
REMARK        NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT       Contact: MGC help desk
              Email: cgabbs-remail.nih.gov
              Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
              cDNA Library Preparation: Michael J. Brownstein (NECRI) & Shiraki
              Toshiyuki and Piero Carninci (RIKEN)
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Sequencing Group at the Stanford Human Genome
              Center, Stanford University School of Medicine, Stanford, CA 94305
              Web site: http://www.shgc.stanford.edu
              Contact: (Dickson, Mark) mcdpaxil.stanford.edu
              Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
              R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK plate: 77 Row: n Column: 20
This clone has the following problem: frame shifted.

FEATURES             source
    source
    1..1747
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:5311213"
    /tissue_type="Brain, hypothalamus"
    /clone_lib="NIH MGC_96"
    /lab_host="DH10B"
    /note="vector: pBluescript"

ORIGIN
Query Match      44.6%; Score 637.2; DB 11; Length 1747;
Best Local Similarity 99.4%; Pred. No. 2.6e-170;
Matches 650; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy      1 ATGGCTTCCCAAGAGAACTTCAGGGCTTTGGCTGCAACCAATCAAGCCCAATGACT 60
Db      429 ATGGCTTCCCAAGAGAACTTCAGGGCTTTGGCTGCAACCAATCAAGCCCAATGACT 488
Qy      61 GAGATGGAGAAATCACTTTTCAGTAATGTTCAGTATGTGATATCTCTGTGAAGAA 120
Db      489 GAGATGGAGAAATCACTTTTCAGTAATGTTCAGTATGTGATATCTCTGTGAAGAA 548
Qy      121 CAGGGAGTGAAGAACATTTTGTGTAATGGCACAACAGAGAGAGCCCTGCTCCCTGAGCGTC 180

```

```

Db      549 CAGGAGTGAAGAACATTTTGTGATCGCACACAGAGAGAGCCCTGTCCCTGAGCGTC 608
Qy      181 TCAGAGCTCCCAAGTTTCAGAGAGAGTGGGTGACAAAGGAGAGACAGCTGATCAG 240
Db      609 TCAGAGCTCCCAAGTTTCAGAGAGAGTGGGTGACAAAGGAGAGACAGCTGATCAG 668
Qy      241 GTGATAATTCAGTAGGAGCACTGAGCTTGAAGGAGTCCACAGAACTGSCCAACATGCA 300
Db      669 GTGATAATTCAGTAGGAGCACTGAGCTTGAAGGAGTCCACAGAACTGSCCAACATGCA 728
Qy      301 GCAGAAATAGGAGCTGATGGCATCGCTGTCTATTGACCCGTTCTTCTCAAGCAATGGACC 360
Db      729 GCAGAAATAGGAGCTGATGGCATCGCTGTCTATTGACCCGTTCTTCTCAAGCAATGGACC 788
Qy      361 AAAGATATCTGATTAATTTCTTAAAGGAAGTGGCTGCTGCCGCCCTGCCCCGCAATTT 420
Db      789 AAAGATATCTGATTAATTTCTTAAAGGAAGTGGCTGCTGCCGCCCTGCCCCGCAATTT 848
Qy      421 TATTACTATCAATTCCTGCTTGACAGGGTAAAGATTCTGCTGAGGAGTTGTGGAT 480
Db      849 TATTACTATCAATTCCTGCTTGACAGGGTAAAGATTCTGCTGAGGAGTTGTGGAT 908
Qy      481 GGGATTCTGGATAAGATCCCACTTCCAAAGGGCTGAAATTCAGTGATACAGATCTCTTA 540
Db      909 GGGATTCTGGATAAGATCCCACTTCCAAAGGGCTGAAATTCAGTGATACAGATCTCTTA 968
Qy      541 GACTTGGGCAATGTGTGATCAGATCCCAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db      969 GACTTGGGCAATGTGTGATCAGATCCCAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1028
Qy      601 GATGAGCACTGTTCAGTGTCTCTGCTGATGGAGCAACTGGAGCAGTGGGCACT 654
Db      1029 GATGAGCACTGTTCAGTGTCTCTGCTGATGGAGCAACTGGAGCAGTGGGCACT 1081

RESULT 8
CA423436      709 bp      mRNA      linear      EST 07-NOV-2002
LOCUS         UI-H-FEI-bec-g-13-0-UI.s1 NCI_CGAP_FEI Homo sapiens cDNA clone
DEFINITION    UI-H-FEI-bec-g-13-0-UI 3', mRNA sequence.
ACCESSION     CA423436
VERSION       CA423436.1 GI:24786162
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1. (bases 1 to 709)
AUTHORS       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL       Unpublished (1997)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-remail.nih.gov
              Tissue Procurement: James Martin
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Clone distribution information can be obtained
              from Dr. M. Bento Soares, bento-soares@uiowa.edu
              Seq primer: M13 FORWARD
              POLYA=Yes.
              Location/Qualifiers
                  1..709
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="UI-H-FEI-bec-g-13-0-UI"
                  /tissue_type="Cell lines"
                  /dev_stage="Adult"
                  /lab_host="DH10B (Life Technologies)"
                  /clone_lib="NCI_CGAP_FEI"
                  /note="Organ: Chondrosarcoma; Vector: pT7T3-Pac"

```

(Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP FEI is a normalized cDNA library derived from a pool of mRNA obtained from 3 cell lines from grade II chondrosarcoma tissues. The library was constructed according to Bonaleo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dr primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGTACGGAC. The cell lines were provided by Dr James Martin from the University of Iowa.
TAG TISSUE=Human grade 2 chondrosarcoma cell line pool
TAG LIB=UT-H-FEI
TAG_SEQ=CGTACGGAC"

ORIGIN

Query Match	42.6%	Score 608.2;	DB 14;	Length 709;
Best Local Similarity	98.8%	Pred. No. 3.3e-162;		
Matches 655;	Conservative 0;	Mismatches 3;	Indels 5;	Gaps 4;

Qy	341	TCCTCTCAAGCCATGACCAAGATATCCGATTAATTTCTTAAGGAAGTGGCTGCTG	400
Db	670	TCCTCTCAAGCCATG--ACCAAGATATCCG--TAATTTCTTAAGGAAGTGG--TGCTG	615
Qy	401	CGGCCCCGCTCCCTCCATTTTATTTACTATCATCTCTCGCTCTGAC--AGGGGTAAAGATT	459
Db	614	CGGCCCCGCTCCCTCCATTTTATTTACTATCATCTCTCGCTCTGAC--AGGGGTAAAGATT	555
Qy	460	CGTCTCAGGAGTTGTTGGATGGGATTTCTGGATAAGATCCCACTTCCAAAGGGCTGAAA	519
Db	554	CGTCTCAGGAGTTGTTGGATGGGATTTCTGGATAAGATCCCACTTCCAAAGGGCTGAAA	495
Qy	520	TTCACTGATACAGATCTTTAGACTTCGGGAAATGTTGTGATCAGATCCGACGACACAG	579
Db	494	TTCACTGATACAGATCTTTAGACTTCGGGAAATGTTGTGATCAGATCCGACGACACAG	435
Qy	580	TTTCTTTCTTTTGGGCTGGATGAGCACTGTTGAGTCTCTGCTGATGGGAGCACT	639
Db	434	TTTCTTTCTTTTGGGCTGGATGAGCACTGTTGAGTCTCTGCTGATGGGAGCACT	375
Qy	640	GGACGATGGGCGATTTTGTATCCAGAGATTATCAACTTTGTTGTCAACTAGGTTTG	699
Db	374	GGACGATGGGCGATTTTGTATCCAGAGATTATCAACTTTGTTGTCAACTAGGTTTG	315
Qy	700	GAGTGTACAGACCAACCCATCATGCTCTGCTCTGGGATTCGAATGGGCCACCC	759
Db	314	GAGTGTACAGACCAACCCATCATGCTCTGCTCTGGGATTCGAATGGGCCACCC	255
Qy	760	GGCTTCCACTGCAGAAAGCTCCAGGGAGTTTACTGATAGTCTGAAGCTAACTGAAGA	819
Db	254	GGCTTCCACTGCAGAAAGCTCCAGGGAGTTTACTGATAGTCTGAAGCTAACTGAAGA	195
Qy	820	GCCTGGATTTCCCTTTCTACTGATTTAAAGGATGGAACTTGGAACTGGTAGCTAGT	879
Db	194	GCCTGGATTTCCCTTTCTACTGATTTAAAGGATGGAACTTGGAACTGGTAGCTAGT	135
Qy	880	GCCTCTCTATCAATCAGGGTTGCACTTGGACATATCTACCTTAAATAGTGCATT	939
Db	134	GCCTCTCTATCAATCAGGGTTGCACTTGGACATATCTACCTTAAATAGTGCATT	75
Qy	940	TTTTCTCAGGAATTTAGATGAATCAATTAATCTCTAGCAAAATGAATCTCAAA	999
Db	74	TTTTCTCAGGAATTTAGATGAATCAATTAATCTCTAGCAAAATGAATCTCAAA	15
Qy	1000	TAA 1002	
Db	14	AAA 12	

RESULT 3	AK088859	1403 bp	linear	HTC 20-SEP-2003
LOCUS	Mus musculus 2 days neonate thymic cells cDNA, RIKEN			
DEFINITION	full-length enriched library, clone:E430029E06			
ACCESSION	AK088859	GI:26353975		
VERSION	HTC; CAP trapper.			
KEYWORDS	Mus musculus (house mouse)			
SOURCE	Mus musculus			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 10349636			
REFERENCE	2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159			
REFERENCE	3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,K., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,T., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861			
REFERENCE	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)			
REFERENCE	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 1403) Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Benio,H., Carninci,P., Fukuda,S., Furumori,M., Hanagaki,T., Hara,A., Hashizume,M., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Kondo,H., Konda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saigo,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sobabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akita,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokotama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gs.c.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)			
REFERENCE	6 (bases 1 to 1403) Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Benio,H., Carninci,P., Fukuda,S., Furumori,M., Hanagaki,T., Hara,A., Hashizume,M., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Kondo,H., Konda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saigo,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sobabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akita,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokotama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gs.c.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)			
COMMENT	cDNA library was prepared and sequenced in Mouse Genome			

602 TGTCTTTAG 610

Db

RESULT 12

AV728536

LOCUS

DEFINITION

AV728536 HTC Homo sapiens cDNA clone HTCB011 5', mRNA sequence.

ACCESSION

AV728536

VERSION

AV728536.1 GI:10837957

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 753)

Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, X., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z. and Han, Z.

TITLE

Homo sapiens CDNA HTC clones

JOURNAL

Unpublished (2000)

COMMENT

Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shouyang Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China

Tel: 86-21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzgehc@sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

source

1..753

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="HTCB011"

/tissue type="Hypothalamus"

/dev stage="Adult"

/lab_host="SOLR"

/clone_lib="HTC"

/note="Vector: pBluescript sk(-); Site 1: EcoRI; Site 2: XhoI"

ORIGIN

Query Match 37.6%; Score 537.2; DB 9; Length 753;

Best Local Similarity 95.0%; Pred. No. 6.7e-142;

Matches 587; Conservative 0; Mismatches 28; Indels 3; Gaps 3;

667 GATTATCAACTTTGTTGCAAACTAGGTTTGGAGTGTACAGACAAAGCCATCAACA 726

90 GTTTCCTCACTTTCTTTTCCGCTAGGTTTGGAGTGTACAGACAAAGCCATCAACA 149

727 CTCGTCTCTGGATTCATATGGGCCACCCCGGCTTCCACTGAGAAAGCTCCAGGG 786

150 CTCGTCTCTGGATTCATATGGGCCACCCCGGCTTCCACTGAGAAAGCTCCAGGG 209

787 AGTTTACTGATAGTGTGAGCTTAACTGAGAGCTGAGTGTCTCTATCAATCAGGTTTGCAC 846

210 AGTTTACTGATAGTGTGAGCTTAACTGAGAGCTGAGTGTCTCTATCAATCAGGTTTGCAC 269

847 TAAAGATGAACTTGGAGCTGTTGAGTGTCTCTATCAATCAGGTTTGCAC 906

270 TAAAGATGAACTTGGAGCTGTTGAGTGTCTCTATCAATCAGGTTTGCAC 329

907 CTTGAGACATATCTACCTTAATAGTGTCTTTTCTCAGGGAATTTAGATGAACCTT 966

330 CTTGAGACATATCTACCTTAATAGTGTCTTTTCTCAGGGAATTTAGATGAACCTT 389

967 GAATAAATCTCTCTAGCAATGAAATCTCACATAAGCAATGAGTACTTTTGTGAGCC 1026

390 GAATAAATCTCTCTAGCAATGAAATCTCACATAAGCAATGAGTACTTTTGTGAGCC 449

1027 TTAATAAGCTTATTTGTGAGGGGCAAACTCTAGAGCTCAACTCTCAGTCACTTC 1086

(PNA) oligos
(N-terminal)-biotin-GTC-CAC-CCG-AAG-CTT-G-(C-terminal) and
(N-terminal)-biotin-C(G/C)-T-GAA-GTT-CTC-AGG-A-(C-terminal)
- Synthesized cDNA was digested with SfiI and
size-selected on a 1% agarose gel (>800bp). Large-scale
sequencing of the library was performed by the NIH
Intramural Sequencing Center (NISC;
Http://www.nisc.nih.gov/).

ORIGIN	
Query Match	37.2%; Score 531.4; DB 13; Length 643;
Best Local Similarity	99.8%; Pred. No. 2.8e-140;
Matches 532; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1 ATGGCCCTCCCAAGAGAACTTCAGGCTCTTGGCTGCACCCATCACGCCAATGACT 60
DB	111 ATGGCCCTCCCAAGAGAACTTCAGGCTCTTGGCTGCACCCATCACGCCAATGACT 170
QY	61 GAGAAATGAGAAATCACTTTTCAGTAAATGGTCAATGATGGATTAATCTTGTGAAGAA 120
DB	171 GAGAAATGAGAAATCACTTTTCAGTAAATGGTCAATGATGGATTAATCTTGTGAAGAA 230
QY	121 CAGGAGTGAAGACATTTTGTGATGSCACACAGAGAGAGCCTGTCTCCCTGAGGCTC 180
DB	231 CAGGAGTGAAGACATTTTGTGATGSCACACAGAGAGAGCCTGTCTCCCTGAGGCTC 290
QY	181 TCAGAGGCTCCCGAGTTCAGAGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 240
DB	291 TCAGAGGCTCCCGAGTTCAGAGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 350
QY	241 GTGTAATTCAGTAGAGCTGAGCTTGAAGAGTCACAGGAGTGGCTGACCATGCA 300
DB	351 GTGTAATTCAGTAGAGCTGAGCTTGAAGAGTCACAGGAGTGGCTGACCATGCA 410
QY	301 GCAGAAATAGGAGCTGAGTGCATGCTCTATTCACCGTCTCTCTCAAGCCATGAGCC 360
DB	411 GCAGAAATAGGAGCTGAGTGCATGCTCTATTCACCGTCTCTCTCAAGCCATGAGCC 470
QY	361 AAAGATATCTTGATTAATTTCTTAAGGAAGTGGCTGTGGCGCCCTGCCCTGCCATTT 420
DB	471 AAAGATATCTTGATTAATTTCTTAAGGAAGTGGCTGTGGCGCCCTGCCCTGCCATTT 530
QY	421 TATTACTATCATCTTCCTGCTGACAGGGTGAAGATTCGTGCTGAGAGTGTGTGGAT 480
DB	531 TATTACTATCATCTTCCTGCTGACAGGGTGAAGATTCGTGCTGAGAGTGTGTGGAT 590
QY	481 GGGATTCGGTAAGATCCCACTTCACAGGCTGAAATTCAGTGATACAGA 533
DB	591 GGGATTCGGTAAGATCCCACTTCACAGGCTGAAATTCAGTGATACAGA 643

RESULT 14
CB999084
LOCUS
DEFINITION
AGENCY 13631478 NIH_MGC 186 Homo sapiens cDNA clone
IMAGE:30325020 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 893)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM130 row: c column: 13
High quality sequence stop: 560.
Location/Qualifiers
1-893
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30325020"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC 186"
/note="Organ: Pooled-Skin; Vector: pDNR-LIB; Site 1: SfiI
(ggccattatggcc); Site 2: SfiI (ggcgactggcc); Library is
oligo-dt primed and directionally cloned. cDNA was
prepared from a pooled samples of tissues from Skin,
meninges, duramater, pia matter and choroid plexus.
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3'
sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, or T). Average
insert size 1.47 kb (range 0.50-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library"

ORIGIN	
Query Match	37.2%; Score 531; DB 14; Length 893;
Best Local Similarity	99.1%; Pred. No. 4.3e-140;
Matches 534; Conservative	0; Mismatches 5; Indels 0; Gaps 0;
QY	1 ATGGCCCTCCCAAGAGAACTTCAGGCTCTTGGCTGCACCCATCACGCCAATGACT 60
DB	106 ATGGCCCTCCCAAGAGAACTTCAGGCTCTTGGCTGCACCCATCACGCCAATGACT 165
QY	61 GAGAAATGAGAAATCACTTTTCAGTAAATGGTCAATGATGGATTAATCTTGTGAAGAA 120
DB	166 GAGAAATGAGAAATCACTTTTCAGTAAATGGTCAATGATGGATTAATCTTGTGAAGAA 225
QY	121 CAGGAGTGAAGAAATTTTGTGAATGSCACACAGAGAGAGCCTGTCTCCCTGAGGCTC 180
DB	226 CAGGAGTGAAGAAATTTTGTGAATGSCACACAGAGAGAGCCTGTCTCCCTGAGGCTC 285
QY	181 TCAGAGCCTGCCAGGTTGCAGAGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 240
DB	286 TCAGAGCCTGCCAGGTTGCAGAGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 345
QY	241 GTGTAATTCAGTAGAGCTGAGCTTGAAGAGTGCACAGGAACTGGCCCAACATGCA 300
DB	346 GTGTAATTCAGTAGAGCTGAGCTTGAAGAGTGCACAGGAACTGGCCCAACATGCA 405
QY	301 GCAGAAATAGGAGCTGAGTGCATGCTGTATTCACCGTCTCTCTCAAGCCATGAGCC 360
DB	406 GCAGAAATAGGAGCTGAGTGCATGCTGTATTCACCGTCTCTCTCAAGCCATGAGCC 465
QY	361 AAAGATATCTTGATTAATTTCTTAAGGAAGTGGCTGTGGCGCCCTGCCCTGCCATTT 420
DB	466 AAAGATATCTTGATTAATTTCTTAAGGAAGTGGCTGTGGCGCCCTGCCCTGCCATTT 525
QY	421 TATTACTATCATCTTCCTGCTGACAGGGTGAAGATTCGTGCTGAGAGTGTGTGGAT 480
DB	526 TATTACTATCATCTTCCTGCTGACAGGGTGAAGATTCGTGCTGAGAGTGTGTGGAT 585
QY	481 GGGATTCGGTAAGATCCCACTTCACAGGCTGAAATTCAGTGATACAGATCTCTT 539
DB	586 GGGATTCGGTAAGATCCCACTTCACAGGCTGAAATTCAGTGATACAGATCTCTT 644

RESULT 15
CB985894
LOCUS
DEFINITION
AGENCY 13631478 NIH_MGC 186 Homo sapiens cDNA clone
IMAGE:30325020 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 893)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

DEFINITION AGENCOURT 13458351 NIH_MGC_184 Homo sapiens cDNA clone
 IMAGE:30325735 5', mRNA sequence.
 ACCESSION CB985894
 VERSION CB985894.1 GI:30280418
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 723)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgsabbs-remail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNL at:
<http://image.llnl.gov>
 Plate: NDCM132 row: a column: 08
 High quality sequence stop: 558.
 Location/Qualifiers
 1..723
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30325735"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIH_MGC_184"
 /note="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site 1:
 SfiI (ggccattatggcc); Site 2: SfiI (ggccgctcggcc);
 Library is oligo-dT primed and directionally cloned. cDNA
 was prepared from a glandular pool of tissues from thyroid,
 parathyroid, adrenal, cortex and pineal gland. 5' and 3'
 adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGAGGCGCGCATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.38
 kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC library."

Query March 36.9%; Score 528; DB 14; Length 723;
 Best Local Similarity 98.8%; Pred. No. 2.8e-139;
 Matches 553; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY 1 ATGGCCTTCCCAAGAGAAGAACTTCAGGGTCTTGGCTGCAACCATCAGCCCAATGACT 60
 |||||
 Db 108 ATGGCCTTCCCAAGAGAAGAACTTCAGGGTCTTGGCTGCAACCATCAGCCCAATGACT 167
 |||||

QY 61 GAGAAATGGAGAAATCACTTTTCAGTAATGGTCAGTATGGATTATCTTGTGAAGAA 120
 |||||
 Db 168 GAGAAATGGAGAAATCACTTTTCAGTAATGGTCAGTATGGATTATCTTGTGAAGAA 227
 |||||

QY 121 CAGGGAGTGAAGAACATTTTGTGATGGCAACACAGAGAGAGCCCTGTCCCTGAGGCTC 180
 |||||
 Db 228 CAGGGAGTGAAGAACATTTTGTGATGGCAACACAGAGAGAGCCCTGTCCCTGAGGCTC 287
 |||||

QY 181 TCAGAGCTGCCAGGTTGCAGAGAGTGGGTGACAAAAGGAGGACAAAGCTGGATCAG 240
 |||||
 Db 288 TCAGAGCTGCCAGGTTGCAGAGAGTGGGTGACAAAAGGAGGACAAAGCTGGATCAG 347
 |||||

QY 241 GTGATTAATTCAGCTAGGACACTGAGCTTGAAGAGTCAAGAACTGGCCCAACATGCA 300
 |||||
 Db 348 GTGATTAATTCAGCTAGGACACTGAGCTTGAAGAGTCAAGAACTGGCCCAACATGCA 407
 |||||

QY 301 GCAGAAATAGGAGCTGATGGCTGCTGATTCATTCGACCCGTTCTTCTCAAGCCATGACC 360
 |||||

Db 408 GCAGAAATAGGAGCTGATGGCTGCTGATTCGACCCGTTCTTCTCAAGCCATGACC 467
 |||||

QY 361 AAAGATATCCTGATTAAATTTCTAAAGGAAGTGGCTGCTGCCGCCCTGCCCTGCCATT 420
 |||||

Db 468 AAAGATATCCTGATTAAATTTCTAAAGGAAGTGGCTGCTGCCGCCCTGCCCTGCCATT 527
 |||||

QY 421 TATTACTATCAGATTCTGCTGCTTGAACAGGGTAAAGATTCGCTGCTGAGGAGTGTGGAT 480
 |||||

Db 528 TATTACTATCAGATTCTGCTGCTTGAACAGGGTAAAGATTCGCTGCTGAGGAGTGTGGAT 587
 |||||

QY 481 GGGATTCTGGATAAGTCCCGACCTT-CCAAGGGCTGAAATTCAGTGTACAGAT-CTCT 538
 |||||

Db 588 GGGATTCTGGATAAGTCCCGACCTTCCAAGGGCTGAAATTCAGTGTACAGAT-CTCT 647
 |||||

QY 539 TAGACTTCGGGCAATGTGTT 558
 |||||

Db 648 TAACTTCGGGCAATGTGTT 667
 |||||

Search completed: May 11, 2004, 11:45:32
 Job time : 4160.87 secs

FEATURES
 source

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:29 ; Search time 109.968 Seconds
(without alignments)
7211.402 Million cell updates/sec

Title: US-09-930-440B-1
Perfect score: 1429
Sequence: 1 atggccttcctccaaagaagaa.....cttttgattcatttcgatg 1429

Scoring table: IDENTITY NUC
Gapop 1C.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1429	100.0	1429	4	US-09-516-143A-5
2	1332.4	93.2	2536	2	US-09-027-013-2
3	1332.4	93.2	2536	3	US-09-244-233-2
4	792	55.4	1923	4	US-09-833-381-1100
5	654	45.8	1127	4	US-09-833-381-1099
6	68.4	4.8	903	4	US-09-543-681A-2696
7	51.4	3.6	927	1	US-09-543-681A-254
8	51.4	3.6	7218	1	US-08-232-463-14
9	44.2	3.1	1368	3	US-08-874-563-5
10	44.2	3.1	1368	3	US-08-577-483-14
11	44.2	3.1	4253	3	US-08-577-483-7
12	43.2	3.0	9897	4	US-08-561-527-10
13	41.8	2.9	4254	2	US-08-443-639-7
14	41.8	2.9	4254	4	US-09-435-380-7
15	41.8	2.9	23439	4	US-08-556-171B-38
16	39.8	2.8	1664976	4	US-08-916-421B-1
17	39.6	2.8	4239	4	US-08-556-171B-333
18	39.2	2.7	832	4	US-09-621-976-2813
19	38.8	2.7	5181	1	US-08-257-073-10
20	38.6	2.7	832	4	US-09-621-976-2813
21	38.6	2.7	540881	4	US-09-750-988-1
22	38.4	2.7	99916	4	US-09-816-095-3
23	38	2.7	1128	4	US-09-710-000-9
24	37.8	2.6	128779	4	US-09-497-855A-38
25	37.6	2.6	1135	4	US-08-936-165A-135
26	37.6	2.6	640681	4	US-09-790-388-1
27	37.2	2.6	128779	4	US-09-497-855A-38

C	28	37	2.6	1219	4	US-08-195-705-1	Sequence 1, Appli
	29	36.8	2.6	6831	2	US-08-609-049A-27	Sequence 27, Appl
	30	36.8	2.6	6831	3	US-09-170-996-27	Sequence 27, Appl
	31	36.8	2.6	8961	4	US-10-204-708-80	Sequence 80, Appl
C	32	36.4	2.5	2614	4	US-09-004-056-1	Sequence 1, Appli
	33	36.4	2.5	5340	4	US-09-627-122-21	Sequence 21, Appl
	34	36.4	2.5	9636	1	US-08-323-170B-1	Sequence 1, Appli
C	35	36.4	2.5	9636	4	US-08-954-441-1	Sequence 1, Appli
	36	36	2.5	5152	4	US-09-833-381-831	Sequence 831, App
	37	36	2.5	5152	4	US-10-204-708-73	Sequence 73, App
	38	35.8	2.5	24979	2	US-08-147-777-3	Sequence 3, Appli
	39	35.8	2.5	24979	3	US-08-452-872-3	Sequence 3, Appli
	40	35.8	2.5	24979	5	PCT-US93-03885-3	Sequence 59, Appl
C	41	35.4	2.5	6156	4	US-10-204-708-59	Sequence 3, Appli
	42	35.4	2.5	65042	4	US-09-784-316-3	Sequence 107, App
	43	35.2	2.5	6405	4	US-08-961-527-107	Sequence 72, Appl
	44	35.2	2.5	8607	4	US-10-204-708-72	Sequence 11, Appl
C	45	35	2.4	9620	3	US-08-952-127-11	

ALIGNMENTS

RESULT 1
US-09-516-143A-5
; Sequence 5, Application US/09516143A
; Patent No. 633482

; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PF050PCT
; CURRENT APPLICATION NUMBER: US/09/516,143A
; CURRENT FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1429
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(690)
US-09-516-143A-5

Query Match 100.0%; Score 1429; DB 4; Length 1429;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGCTTCCCAAGAGAAACTTCAGGCTCTGTGGCTGCAACCATCAAGCCCAATGACT	60
Db	1	ATGGCTTCCCAAGAGAAACTTCAGGCTCTGTGGCTGCAACCATCAAGCCCAATGACT	60
QY	61	GAGAAATGGAATAATCACTTTTCAGTAAATGGTCAGTATGATGATTTCTGTGAAGAA	120
Db	61	GAGAAATGGAATAATCACTTTTCAGTAAATGGTCAGTATGATGATTTCTGTGAAGAA	120
QY	121	CAGGAGTAGGAACAACTTTTGTGAATGGCAACAGGAGAGGCTGTCCCTGAGGCTC	180
Db	121	CAGGAGTAGGAACAACTTTTGTGAATGGCAACAGGAGAGGCTGTCCCTGAGGCTC	180
QY	181	TCAGAGCTGCCAGCTTCAGAGAGAGTGGTGCACAAAGGAGCAAGCTGGATCAG	240
Db	181	TCAGAGCTGCCAGCTTCAGAGAGAGTGGTGCACAAAGGAGCAAGCTGGATCAG	240
QY	241	GTGATAATCACCTAGGAGCACTGAGCTTCAAGGCTCAGAGNACTGGCCCAATGCA	300
Db	241	GTGATAATCACCTAGGAGCACTGAGCTTCAAGGCTCAGAGNACTGGCCCAATGCA	300
QY	301	CGAGAAATAGGAGCTGATGGCATGCTGTCAATGCAACCGTCTTCTCTCAAGCCATGACC	360
Db	301	CGAGAAATAGGAGCTGATGGCATGCTGTCAATGCAACCGTCTTCTCTCAAGCCATGACC	360

361 AAAAGATATCTGATTAATTTCTTAAGGAGATGGCTGTGCGGCGCTGCGCATTT 420
 Db |||||
 361 AAAAGATATCTGATTAATTTCTTAAGGAGATGGCTGTGCGGCGCTGCGCATTT 420
 Qy |||||
 421 TATTACTATCATATCTGCTTGTACAGGGGTAAGATTCGTGCTGAGGATTTGGAT 480
 Db |||||
 421 TATTACTATCATATCTGCTTGTACAGGGGTAAGATTCGTGCTGAGGATTTGGAT 480
 Qy |||||
 481 GGGATTCGGATAGATCCCACTTCCAAAGGCTGAAATTCAGTGATACAGATCTCTTA 540
 Db |||||
 481 GGGATTCGGATAGATCCCACTTCCAAAGGCTGAAATTCAGTGATACAGATCTCTTA 540
 Qy |||||
 541 GACTTCGGGCAATGTTGATCAGATCCGACGACAGATTTGCTTCTCTTTGGGGTG 600
 Db |||||
 541 GACTTCGGGCAATGTTGATCAGATCCGACGACAGATTTGCTTCTCTTTGGGGTG 600
 Qy |||||
 601 GATGAGCAACTGTTGAGTGTCTGCTGATGGGAGCAACTGGAGCAAGTGGGCAATTTGTA 660
 Db |||||
 601 GATGAGCAACTGTTGAGTGTCTGCTGATGGGAGCAACTGGAGCAAGTGGGCAATTTGTA 660
 Qy |||||
 661 TCCAGAGATTTATCAACTTTGTTGTCAAACTAGGTTTGGAGTGCACAGACCAAGCCA 720
 Db |||||
 661 TCCAGAGATTTATCAACTTTGTTGTCAAACTAGGTTTGGAGTGCACAGACCAAGCCA 720
 Qy |||||
 721 TCATGACTCTGCTCTGCGGATTCGAATGGGCGCCACCCGCTTCCACTGCAGAAAGCCT 780
 Db |||||
 721 TCATGACTCTGCTCTGCGGATTCGAATGGGCGCCACCCGCTTCCACTGCAGAAAGCCT 780
 Qy |||||
 781 CCAGGGAGTTTACTGATGCTGAGCTTAAGCTTAAGCTGAAGAGCCTGGATTTCTTTTCA 840
 Db |||||
 781 CCAGGGAGTTTACTGATGCTGAGCTTAAGCTTAAGCTGAAGAGCCTGGATTTCTTTTCA 840
 Qy |||||
 841 CTGATTTAAAGATGGAACCTTGGAGCTGCTAGCTAGTGCCTCTCTATCAAAATCAGGCT 900
 Db |||||
 841 CTGATTTAAAGATGGAACCTTGGAGCTGCTAGCTAGTGCCTCTCTATCAAAATCAGGCT 900
 Qy |||||
 901 TTGCACCTTGACATATCTACCTTAATAGTGCATTTTCTCAGGGATTTTATAGT 960
 Db |||||
 901 TTGCACCTTGACATATCTACCTTAATAGTGCATTTTCTCAGGGATTTTATAGT 960
 Qy |||||
 961 GAACTTGAATAAACCTCTCTAGCAATGAAATCTCACATAAGCACTGAGTACCTTTTG 1020
 Db |||||
 961 GAACTTGAATAAACCTCTCTAGCAATGAAATCTCACATAAGCACTGAGTACCTTTTG 1020
 Qy |||||
 1021 TGAGCCTTAAAGATCTTAATTTTGTGAGAGGGGCAAAACTCTAGAGTCAAACTCTCAG 1080
 Db |||||
 1021 TGAGCCTTAAAGATCTTAATTTTGTGAGAGGGGCAAAACTCTAGAGTCAAACTCTCAG 1080
 Qy |||||
 1081 TCATTCATTTACAGATTTTGTGAGAGAAATTTCTGTTATATGATGAATGGATC 1140
 Db |||||
 1081 TCATTCATTTACAGATTTTGTGAGAGAAATTTCTGTTATATGATGAATGGATC 1140
 Qy |||||
 1141 AAGAGGAAATTTGTAATGATTAATTCATCTGCTTTAGGAGCTCTCATATCTCGGTC 1200
 Db |||||
 1141 AAGAGGAAATTTGTAATGATTAATTCATCTGCTTTAGGAGCTCTCATATCTCGGTC 1200
 Qy |||||
 1201 TCTGTTCTTATCTTATTTAAAGTGTCTTAATTTTAAACCACTATAATGTTCTCAT 1260
 Db |||||
 1201 TCTGTTCTTATCTTATTTAAAGTGTCTTAATTTTAAACCACTATAATGTTCTCAT 1260
 Qy |||||
 1261 TTTTATAATATTTCAATTCGAAATCTAGGAAACTCTGAGCTACTGATTTAGGCGAGGAC 1320
 Db |||||
 1261 TTTTATAATATTTCAATTCGAAATCTAGGAAACTCTGAGCTACTGATTTAGGCGAGGAC 1320
 Qy |||||
 1321 TTTTATACCAAACTGTAACATCTCAACTGTATACAACTCAAAATACACCAAGCTCATTT 1380
 Db |||||
 1321 TTTTATACCAAACTGTAACATCTCAACTGTATACAACTCAAAATACACCAAGCTCATTT 1380
 Qy |||||
 1381 GCGTGTGCTAGTCTAATCTAGATGATGCTTTTGAATTCATTTTGGATG 1429
 Db |||||
 1381 GCGTGTGCTAGTCTAATCTAGATGATGCTTTTGAATTCATTTTGGATG 1429

RESULT 2
 US-09-027-013-2
 ; Sequence 2, Application US/09027013
 ; Patent No. 5962302
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Shah, Purvi
 ; APPLICANT: Corley, Neil C.
 ; TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/027,013
 ; FILING DATE: Herewith
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PP-0462 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-855-0555
 ; TELEFAX: 650-845-4166
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2536 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: BBSNOT07
 ; CLONE: 2125668
 ; US-09-027-013-2

Query Match 93.2%; Score 1332.4; DB 2; Length 2536;
 Best Local Similarity 94.3%; Pred. No. 0;
 Matches 1428; Conservative 0; Mismatches 1; Indels 85; Gaps 1;
 Qy 1 ATGGCGCTTCCAAAGAGAAACTTCAGGCTCTGTGGCTGCAACCATCAGCCCAATGACT 60
 Db 89 ATGGCGCTTCCAAAGAGAAACTTCAGGCTCTGTGGCTGCAACCATCAGCCCAATGACT 145
 Qy 61 GAGATGAGAAATCAACTTTTTCAGTAATTTGTCAGTATGTCATATCTTGTGAAGAA 123
 Db 149 GAGATGAGAAATCAACTTTTTCAGTAATTTGTCAGTATGTCATATCTTGTGAAGAA 208
 Qy 121 CAGGGAGTGAAGAACTTTTTCGTAATGCGCAACAGGAGAGGCTGTCTTGTAGCGTC 180
 Db 209 CAGGGAGTGAAGAACTTTTTCGTAATGCGCAACAGGAGAGGCTGTCTTGTAGCGTC 268
 Qy 181 TCAGAGCTGCGCCAGTTCAGAGGAGTGGGTGACAAAGGAGGAGGAGGAGGAGTATGAT 240
 Db 269 TCAGAGCTGCGCCAGTTCAGAGGAGTGGGTGACAAAGGAGGAGGAGGAGGAGTATGAT 328
 Qy 241 GTGATAATTCAGTATGAGGAGTTCAGTATGAGGAGTTCAGGAACTGCGCCCAATGCA 300
 Db 329 GTGATAATTCAGTATGAGGAGTTCAGTATGAGGAGTTCAGGAACTGCGCCCAATGCA 388
 Qy 301 GCAGAAATAGGAGCTGATGGCATCGTGTCAATTGCAACCGTTCTTCTCAAGCCATGAGC 360

Db GCAGAAATAGAGCTGATGGCATCGCTGCTCATGACCGCTTCTCTCAAGCCATGGACC 448
QY AAAGATATCTGATTAATTTCTAAAGAAATGGCTGCTGCGGCCCTGCGCTGCATTT 420
Db AAAGATATCTGATTAATTTCTAAAGAAATGGCTGCTGCGGCCCTGCGCTGCATTT 508
QY TATTACTATCATCATTTCTGCTGCTGACAGGGTAAAGATTGCTGAGGAGTTGTGGAT 480
Db TATTACTATCATCATTTCTGCTGCTGACAGGGTAAAGATTGCTGAGGAGTTGTGGAT 568
QY GGGATTTCTGATAAGATCCCACTTCCAAAGGCTGAAATTCAGTGATACAGATCTCTTA 540
Db GGGATTTCTGATAAGATCCCACTTCCAAAGGCTGAAATTCAGTGATACAGATCTCTTA 628
QY GACTTCGGGCAATGTTGTGATCAGAAATCGCCAGCAAGTTGCTTCTCTTTTGGGGTG 600
Db GACTTCGGGCAATGTTGTGATCAGAAATCGCCAGCAAGTTGCTTCTCTTTTGGGGTG 688
QY GATGAGCAATTTGAGTGTCTGCTGATGGAGCAACTGGAGCAGTGGGAGTACCTAT 748
Db GATGAGCAATTTGAGTGTCTGCTGATGGAGCAACTGGAGCAGTGGGAGTACCTAT 650
QY AACTACTCTGGAAAAAGACAAACAGATGTTGGAGGCTTTGACAAAAGGACTTCTCT 808
Db AACTACTCTGGAAAAAGACAAACAGATGTTGGAGGCTTTGACAAAAGGACTTCTCT 651
QY TTAGCCCTGAATTCAGTTTCTTATCCAGATTTATCAACTTGTGTCAAACTAGGT 695
Db TTAGCCCTGAATTCAGTTTCTTATCCAGATTTATCAACTTGTGTCAAACTAGGT 868
QY TTTGGAGTGTACAGACCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 755
Db TTTGGAGTGTACAGACCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 928
QY CCCCCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 815
Db CCCCCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 929
QY AAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 875
Db AAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 988
QY TATGCTCTCTATCAATCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1048
Db TATGCTCTCTATCAATCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 875
QY TAGTCTCTCTATCAATCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 935
Db TAGTCTCTCTATCAATCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1048
QY ATTTTCTCTGAGGAACTTTAGATGAATGAATGAATGAATGAATGAATGAATGAATGTC 995
Db ATTTTCTCTGAGGAACTTTAGATGAATGAATGAATGAATGAATGAATGAATGAATGTC 1109
QY ACAATAAGCAATTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
Db ACAATAAGCAATTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1168
QY AACTCTAGGAGTCAAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1228
Db AACTCTAGGAGTCAAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
QY AACTCTAGGAGTCAAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1115
Db AACTCTAGGAGTCAAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1228
QY CTGTTTATATGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGTC 1175
Db CTGTTTATATGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGTC 1348
QY TTTAGGAGTCTCATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
Db TTTAGGAGTCTCATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1349
QY TTAACCACTATAATATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1408
Db TTAACCACTATAATATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1235
QY TGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
Db TGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

Db 1469 TGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1528
QY CAACCTCAAAATACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
Db CAACCTCAAAATACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1588
QY AATTCAATTCGATG 1429
Db AATTCAATTCGATG 1602

RESULT 3

US-09-244-233-2

; Sequence 2, Application US/09244233

; Patent No. 6030824

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Shah, Purvi

APPLICANT: Corley, Neil C.

TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: Fast-Seq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09244,233

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/027,013

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0462 US

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2536 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: BRSTNCT07

CLONE: 2125668

US-09-244-233-2

Query Match

Best Local Similarity 93.2%; Score 1332.4; DB 3; Length 2536;

Matches 1428; Conservative C; Mismatches 1; Indels 85; Gaps 1;

QY 1 ATGGCTTCCCAAGAGAACTTCAGGGTCTTGGCTGCAACCATCACGCCAATGACT 60
Db 89 ATGGCTTCCCAAGAGAACTTCAGGGTCTTGGCTGCAACCATCACGCCAATGACT 148
QY 61 GAGATGAGAAATCACTTTTCAGTAATTCAGTATGCGATTCATCTTGAAAGAA 120
Db 149 GAGATGAGAAATCACTTTTCAGTAATTCAGTATGCGATTCATCTTGAAAGAA 208
QY 121 CAGGAGTGAAGAACTTTTGTGATGCGCAACAGAGAGGCTGCTCCCTGAGCCTC 180
Db 209 CAGGAGTGAAGAACTTTTGTGATGCGCAACAGAGAGGCTGCTCCCTGAGCCTC 268


```
QY 865 AAGCTGCTAGCTAGTGGCTCTCTATCAATCAGGGTTGGCCTTGGACATAATCTACC 924
Db 1503 AAGCTGCTAGCTAGTGGCTCTCTATCAATCAGGGTTGGCCTTGGACATAATCTACC 1444
QY 925 TTAATAGTGCATTTTCTCAGGAAATTTAGATGAACCTTGATTAACCTCTCTAGCA 984
Db 1443 TTAATAGTGCATTTTCTCAGGAAATTTAGATGAACCTTGATTAACCTCTCTAGCA 1384
QY 985 AATGAATCTCACAAATAGCAATCAGGTACCTTTTGTGAGCCTTAAAGTCTTAATTTG 1044
Db 1383 AATGAATCTCACAAATAGCAATCAGGTACCTTTTGTGAGCCTTAAAGTCTTAATTTG 1324
QY 1045 TGAAGGGCAAAAACCTCTAGGAGTCAACAATCTCAGTCATTCATTTTCACAGATTTTGTG 1104
Db 1323 TGAAGGGCAAAAACCTCTAGGAGTCAACAATCTCAGTCATTCATTTTCACAGATTTTGTG 1264
QY 1105 TGGAGAAATTTCTCTTATATGGATGAATGAATCAAGAGAAATTTGTAATGATTA 1164
Db 1263 TGGAGAAATTTCTCTTATATGGATGAATGAATCAAGAGAAATTTGTAATGATTA 1204
QY 1165 TCCCATCTGCTTAGGAGCTCTCATATCTCGGTCTCTGCTCTCTATCTTAATCTTAATTA 1224
Db 1203 TCCCATCTGCTTAGGAGCTCTCATATCTCGGTCTCTGCTCTCTATCTTAATCTTAATTA 1144
QY 1225 GTTCTCTPAATTTAAACCCCTATATATGCTCTCAATTTTAAATAATATTCATTTTGAATC 1284
Db 1143 GTTCTCTPAATTTAAACCCCTATATATGCTCTCAATTTTAAATAATATTCATTTTGAATC 1084
QY 1285 TAGGAACCTCTGAGCTACTGCATTTAGGAGGCACTTTAATACAACTCTAATGTC 1344
Db 1083 TAGGAACCTCTGAGCTACTGCATTTAGGAGGCACTTTAATACAACTCTAATGTC 1024
QY 1345 TCACTGTATACAACTCAAAATACACAGCTCATTTGGCTGCTCAGCTAATCTAGAT 1404
Db 1023 TCACTGTATACAACTCAAAATACACAGCTCATTTGGCTGCTCAGCTAATCTAGAT 964
QY 1405 -GAGTCTTTGAAATTCATTTGGATG 1429
Db 963 GGGATGCTTTGAATTCATTTGGATG 938
```

RESULT 5

```
US-09-833-381-1099
; Sequence 1099, Application US/39833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1099
; LENGTH: 1127
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-1099
```

```
Query Match 45.8%; Score 654; DB 4; Length 1127;
Best Local Similarity 100.0%; Pred. No. 8.2e-186;
Matches 654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGGCCCTCCCAAGAGAACTTCAGGCTCTGTGGCTGCAACCATCAGGCAATGACT 60
Db 199 ATGGCCCTCCCAAGAGAACTTCAGGCTCTGTGGCTGCAACCATCAGGCAATGACT 258
QY 61 GAGAAATGAGAAATCAACTTTTCAGTAATTTGTCAGTATGTCGATTTATCTTGTGAAGAA 120
Db 259 GAGAAATGAGAAATCAACTTTTCAGTAATTTGTCAGTATGTCGATTTATCTTGTGAAGAA 318
```

```
QY 121 CAGGAGCTGAAGAACATTTTCTGAATGGCAACAACAGAGAGAGGCTCTCTCTAGCGTC 180
Db 319 CAGGAGCTGAAGAACATTTTCTGAATGGCAACAACAGAGAGAGGCTCTCTCTAGCGTC 378
QY 181 TCAGAGCCTCGCCACAGTTTGCAGAGGAGTGGGTGACAAAGAGGAGCAAGCTGGATCAG 240
Db 379 TCAGAGCCTCGCCACAGTTTGCAGAGGAGTGGGTGACAAAGAGGAGCAAGCTGGATCAG 438
QY 241 GTGATAATTCAGTATAGGAGCACTGAGCTTGAAGAGTCAACAGAACTGSCCCCAATGCA 300
Db 439 GTGATAATTCAGTATAGGAGCACTGAGCTTGAAGAGTCAACAGAACTGSCCCCAATGCA 498
QY 301 GCAGAAATAGAGCTGATGSCATCTGCTCAATTCACCGTCTTCTCTAGAGCATGGACC 360
Db 499 GCAGAAATAGAGCTGATGSCATCTGCTCAATTCACCGTCTTCTCTAGAGCATGGACC 558
QY 361 AAAGATATCTGATTAATTTCTTAAAGGAAGTGGCTGCTGCGCCCTGCGCTGCCATTT 420
Db 559 AAAGATATCTGATTAATTTCTTAAAGGAAGTGGCTGCTGCGCCCTGCGCTGCCATTT 618
QY 421 TATTACTATCATTTCTGCTTGAAGAGTGAAGATTCAGTATGAGAGTGTGGAT 480
Db 619 TATTACTATCATTTCTGCTTGAAGAGTGAAGATTCAGTATGAGAGTGTGGAT 678
QY 481 GGGATTTCTGATAAGATCCCACTTCCAGSGCTGAAATTCAGTATGATACATCTCTTA 540
Db 679 GGGATTTCTGATAAGATCCCACTTCCAGSGCTGAAATTCAGTATGATACATCTCTTA 738
QY 541 GACTTGGGCAATGTGTGATCAGAAATCGCCAGCAACAGTTTGTCTTCTTTTGGGCTG 600
Db 739 GACTTGGGCAATGTGTGATCAGAAATCGCCAGCAACAGTTTGTCTTCTTTTGGGCTG 798
QY 501 GATCAGCAACTGTGTGAGTGTCTGCTGATGGAGCAACTGAGCAGTGGGCACT 654
Db 799 GATCAGCAACTGTGTGAGTGTCTGCTGATGGAGCAACTGAGCAGTGGGCACT 852
```

RESULT 6

```
US-09-543-681A-2696
; Sequence 2696, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2696
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2696
```

```
Query Match 4.8%; Score 68.4; DB 4; Length 903;
Best Local Similarity 49.3%; Pred. No. 2.1e-10;
Matches 267; Conservative 0; Mismatches 266; Indels 9; Gaps 3;
```

```
QY 13 AAGAAGAAATTCAGAGCTCTGTGCTGCAACATCAGCGCAATGACTGAGATGAGAA 72
Db 10 ATGAACAAATCTCTGAGCTGATTCGCGCTCTCATACTCTTTTCTGCGGAGTAGT 69
QY 73 ATCAACTTTTCAGTAAATTTGCTCAGTATGTSATTTATCTTGTGAAGAACAGGAGTGAAG 132
Db 70 GTTAATTTACCTGTTATTG---ATGATATGCTAAGCACTAATTAAGACTGGGTAACA 126
QY 133 AACATTTTGTGAATGGCAACAACAGAGAGGCTGTCCCTGAGCGCTCTCAGAGCGTGGC 192
Db 127 GGGGCTTATGCTCTAGTACTACTGCTGAAGGTATACATTTGCTCGGTTGAAGAGCGTAAA 186
```


QY 252 CGTAGGACCTGAGCTTGAAGAGTCACAGGAACGGCCCAACATGCAGCAGAAATAGG 311
D 1142 RRR 1083
QY 312 AGCTGATGGCATCGCTGTCATTGACCGCTCTTCTCAAGCCATGACCAAGATATCT 371
D 1082 RRR 1023
QY 372 G 372
D 1022 G 1022

RESULT 9

US-08-874-563-5/c

; Sequence 5, Application US/08874563

; Patent No. 6002071

; GENERAL INFORMATION:

; APPLICANT: CHAPPELL, JOSEPH

; APPLICANT: NEWMAN, JEFFREY D.

; APPLICANT: YIN, SHAOHUI

; TITLE OF INVENTION: TRANSCRIPTIONAL SILENCING

; TITLE OF INVENTION: ELEMENTS AND THEIR BINDING FACTORS

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESS: Clark & Elbing LLP

; STREET: 176 Federal Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02110

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/874,563

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/020,087

; FILING DATE: June 13, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Paul T. Clark, Esq.

; REGISTRATION NUMBER: 30,162

; REFERENCE/DOCKET NUMBER: 07678/008001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-428-0200

; TELEFAX: 617-428-7045

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1368 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

US-08-874-563-5

Query Match 3.1%; Score 44.2; DB 3; Length 1368;

Best Local Similarity 47.1%; Pred. No. 0.0048;

Matches 136; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1007 TGAGGTACCTTTGTGAGCCTTAAAGTCTTATTTGTGAAGGGCAAAACTCTAGGA 1066

D 437 TGACATAGGTATATGATCTTTACGCCCTTTATTTTTCAGAAAAAAGTTCGATGG 378

QY 1067 GTCACAACTCTCAGTCAATTCATTTTCACAGATTTTTTTTGTGAGAAATTTCTGTTTATATG 1126

D 377 TCGCGGCATGCGATACACACACACACATATATATATATATATATATATA 318

QY 1127 GATGAATGGAATCAAGAGGAAATTTGTAATTTTAAAGTGTCTAAATTTTAAACCACTA 1186

D 317 TA-ATA 258

D 317 TAT 258
QY 1187 TCATTATCTCGCTCTCTGGTCTCTTAATCCTATTTTAAAGTTCCTAATTTTAAACCACTA 1246
D 257 TCACAAATTTAGTGTATTTTCTGATTATATATATATATATATATATATATATA 198
QY 1247 TAATATGCTTCATTTTAAATAATATTCATTTGGAAATCTAGGAAATCTC 1295
D 197 AAATAAAACACCATGGGCTTACACCTCGCTTCGTTAGATAAAACGC 149

RESULT 10

US-08-577-483-14/c

; Sequence 14, Application US/08577483

; Patent No. 6100451

; GENERAL INFORMATION:

; APPLICANT: Chappell, Joseph

; APPLICANT: Yin, Shaohui

; APPLICANT: Cornett, Catherine A.G.

; TITLE OF INVENTION: Transcriptional Control Sequences and

; TITLE OF INVENTION: Methods

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/577,483

; FILING DATE: 22-DEC-1995

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Paul T. Clark

; REGISTRATION NUMBER: 32,164

; REFERENCE/DOCKET NUMBER: 07678/003001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1368 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: both

; TOPOLOGY: both

; MOLECULE TYPE: cDNA

US-08-577-483-14

Query Match

Best Local Similarity 47.1%; Pred. No. 0.0048;

Matches 136; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1007 TGAGGTACCTTTGTGAGCCTTAAAGTCTTATTTGTGAAGGGCAAAACTCTAGGA 1066

D 437 TGACATAGGTATATGATCTTTACGCCCTTTATTTTTCAGAAAAAAGTTCGATGG 378

QY 1067 GTCACAACTCTCAGTCAATTCATTTTCACAGATTTTTTTTGTGAGAAATTTCTGTTTATATG 1126

D 377 TCGCGGCATGCGATACACACACACACATATATATATATATATATATATA 318

QY 1127 GATGAATGGAATCAAGAGGAAATTTGTAATTTTAAAGTGTCTAAATTTTAAACCACTA 1186

D 317 TA-ATA 258

QY 1187 TCATTATCTCGGCTCTGTTCCCTAATCTTTTAAAGTGTCTAAATTTTAAACCACTA 1246

Matches 182; Conservative 0; Mismatches 188; Indels 6; Gaps 2;
QY 90 TGGTCAGTATGAGTATCTCTGTGAAGACACAGGAGTGAAGACATTTTGTGAATGG 149
Db |||||
4277 TGTAAATCTCTGGTACATATTTTCATTGACAAAGGTGTAAAGGTATCTATGTAATGG 4218
QY 150 CACAACAGGAGGAGCCCTGCTCCGAGGCTCTCAGAGGCTGCCAGGTTGAGAGGAGTG 209
Db |||||
4217 TCTTCAGGTGAATGATTTTACCAAGTGTAGAAGATCTGAACAAATTAATTGAAGCTGT 4158
QY 210 GGTGACAAAGGAGGAGGACAGCTGATCAGTGTATATTCAGTGGAGGACACTGAGCTT 269
Db |||||
4157 TATGAGGTGCTAAAGTAAATTTTAAAGTATCAACCAATTAATGATGAATGAACAC 4101
QY 270 GAAGAGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 329
Db |||||
4100 GAAGATAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 4041
QY 330 CATTGACCGTTCTTCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 389
Db |||||
4040 TATCCCACTATTTATTTCAA---ATTGCCAGAGTATTCATTCAGGAGGATATGGAATGC 3984
QY 390 AGTGCTGCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCT 449
Db |||||
3983 AATGAGTGAAGTGGTGGTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAGT 3924
QY 450 GGTAAAGATTCGTGCT 465
Db |||||
3923 GGTTCGTTGACTGCT 3308

RESULT 13
US-08-443-639-7/c
Sequence 7, Application US/08443639
Patent No. 5981843
GENERAL INFORMATION:
APPLICANT: Chappell, Joseph
APPLICANT: Yin, Shaohui
APPLICANT: Cornett, Catherine A.G.
TITLE OF INVENTION: Transcriptional Control Sequences and
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443,639
FILING DATE: 18-MAY-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33878
REFERENCE/DOCKET NUMBER: 69-94
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4254 base pairs
TYPE: nucleic acid
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

ANTI-SENSE: NO
ORIGINAL SOURCE: Nicotiana tabacum
FEATURE:
NAME/KEY: CDS
LOCATION: join(1217...1327, 1455...1718, 1806...2182,
LOCATION: 2259
LOCATION: ..2477, 2609...2747, 2903...3148, 3262...3558)
US-08-443-639-7
Query Match 2.9%; Score 41.8; DB 2; Length 4254;
Best Local Similarity 50.2%; Pred. No. 3.047;
Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 1091 CACGATTTTGTGGAGAAATTTCTGTTTATATGATGAATGAATCAAGAGGAAAA 1150
Db |||||
353 CACACATACATA 294
QY 1151 TTGTAATGATTAATTCATCTGTTTAGGAGCTCTCATTTCTCGTCTCTGGTTCTT 1210
Db |||||
293 TAT 234
QY 1211 AATCTATTTTAAAGTGTCTAATTTTAAACCACTATATATATATATATATATATAT 1270
Db |||||
233 ATTAT 174
QY 1271 ATTCATTTGGATCTAGGAAACTC 1295
Db |||||
173 CCTCGCTTCGTTAGATAAAGC 149

RESULT 14
US-09-435-380-7/c
Sequence 7, Application US/09435380
Patent No. 605764
GENERAL INFORMATION:
APPLICANT: Chappell, Joseph
APPLICANT: Yin, Shaohui
APPLICANT: Cornett, Catherine A.G.
TITLE OF INVENTION: Transcriptional Control Sequences and
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/435,380
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/443,639
FILING DATE: 18-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33878
REFERENCE/DOCKET NUMBER: 69-94
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4254 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

```
;
;
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23439 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-08-956-171E-38

Query Match          2.9%; Score 41.8; DB 4; Length 23439;
Best Local Similarity 47.4%; Pred. No. 0.12;
Matches 158; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

QY 13 AAGAGAAACCTTCAGGGTCTTGTGGCTGCAACCATCAAGCCCAATGACTCAGAGATGAGAA 72
DB 8682 AACAAAGATTTAAAGGTTTATATGACGGTTACTCTGTTCTTTTCATGAAATGCTCAA 8623
QY 73 ATCAACTTTTCGTTAATTTGGTCAGTATGTTGGATTATCTTGTGAAAGAACAGGGAGTGAAG 132
DB 8622 GTAAATGAACAGGACITTAACAAATTTCTCAAAATGCCAATTGMAAATGAAAGATTAGAC 8563
QY 133 AACATTTTGTGAATGGCACACACAGGAGAGGCCCTGTCCCTGAGCGTCTCAGAGCGTCCG 192
DB 8562 GGTCTTTAATGTAATGTTAGTCTCAGCGGAAACTTTTATTAA---ATACAGAGCAGAG 8506
QY 193 CAGGTTGCAGAGGAGTGGGTGACAAAGGAGGAGCAAGCTGGATCAGGTGATATTTAC 252
DB 8505 AAGCAAGTTTCAAGGTTGCCAAAGAGGAGGAGTTGGGGATAAAGTGAATTTGATTCTCAA 8446
QY 253 GTAGGAGCACTCAGCTTGAAGGAGTACAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGT 312
DB 8445 GTAGGTTGTTAGATTAAATGAAGCCATTGAAGTGGAAATATATGCCACAGACTCGGT 8386
QY 313 GCTGATGGCATCGGTGTCTATTGCAACCGTTCTTC 345
DB 8385 TATGATGCACTTTCTGCGGTAAACCACTTCTAC 8353

Search completed: May 11, 2004, 07:22:11
Job time : 127.968 secs
```

```
;
;
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23439 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-08-956-171E-38

Query Match          2.9%; Score 41.8; DB 4; Length 4254;
Best Local Similarity 50.2%; Pred. No. 0.047;
Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 1091 CACAGATTTTGTGGAGAAATTTCTGTTTATATGGATGAAATGGAATCAAGAGGAAA 1150
DB 353 CACACATACATTTATATATATATATATATATATATATATATATATATATATATATAT 294
QY 1151 TTGTAATGANTTAATCCATCTGCTTTAGAGCTCTCATATCTCGGTCCTGTTCTCT 1210
DB 293 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 234
QY 1211 AATCCTATTTTAAAGTCTCTAAATTTTAAACCACTATATATATATATATATATAT 1272
DB 233 ATTATATATATATATATATATATATATATATATATATATATATATATATATATAT 174
QY 1271 ATTCAATTTGGAATCTAGAAACTC 1295
DB 173 CCTCGCTTCGTGTAGATAAAAGC 149

RESULT 15
US-08-956-171E-38/c
; Sequence 38, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; GIL H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannin
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknwn>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 11:46:35 ; Search time 662.155 Seconds

(without alignments)
9776.195 Million cell updates/sec

Title: US-09-930-440B-1

Perfect score: 1429

Sequence: 1 atggccttcccaagaagaa.....ctttgaattcatttgatg 1429

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 588172

Minimum DB seq length: 5

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/prodata/2/pubpna/US10E_NEW_PUB.seq.*
- 18: /cgn2_6/prodata/2/pubpna/US10F_NEW_PUB.seq.*
- 19: /cgn2_6/prodata/2/pubpna/US10G_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1429	100.0	1429	9	US-09-984-205-5
2	1429	100.0	1429	9	US-09-930-440B-1
3	1328	92.9	1552	15	US-10-102-524-1751
4	792	55.4	1923	9	US-09-833-381-1100
5	654	45.8	1127	9	US-09-833-381-1099
6	374	26.2	374	15	US-10-102-524-2029
7	312	21.8	400	13	US-10-085-783A-21707
8	312	21.8	400	16	US-10-242-535A-21707
9	149.2	10.4	309	9	US-09-783-590-1834
10	128.2	9.0	472	10	US-09-918-995-8879
11	46.4	3.2	177249	16	US-10-085-117-223
12	45.4	3.2	6117	15	US-10-311-455-998
13	44.8	3.1	5771	15	US-10-311-455-1924
14	44.4	3.1	6109	13	US-10-221-613-33

15	44.4	3.1	6109	15	US-10-311-455-299	Sequence 289, App
16	44.2	3.1	19380	13	US-10-221-613-389	Sequence 389, App
17	44	3.1	8170	15	US-10-240-453-131	Sequence 131, App
18	44	3.1	13574	15	US-10-311-455-1290	Sequence 1290, App
19	43.2	3.0	6361	15	US-10-311-455-1114	Sequence 1114, App
20	43.2	3.0	9897	13	US-10-158-844-10	Sequence 10, Appl
21	43.2	3.0	17848	15	US-10-239-676-28	Sequence 28, Appl
22	43.2	3.0	17848	15	US-10-240-453-38	Sequence 38, Appl
23	43.2	3.0	17848	17	US-10-257-166-58	Sequence 58, Appl
24	42.2	3.0	8991	13	US-10-221-613-256	Sequence 256, App
25	42	2.9	6185	15	US-10-311-455-385	Sequence 385, App
26	42	2.9	3673778	15	US-10-312-841-2	Sequence 2, Appl
27	41.8	2.9	5185	15	US-10-311-455-1007	Sequence 1007, App
28	41.8	2.9	10945	15	US-10-240-453-227	Sequence 227, App
29	41.8	2.9	23439	8	US-08-781-386A-38	Sequence 38, Appl
30	41.8	2.9	23439	13	US-10-329-624-38	Sequence 38934, A
31	41.6	2.9	490	13	US-10-027-632-38934	Sequence 83646, A
32	41.6	2.9	490	13	US-10-027-632-83646	Sequence 83934, A
33	41.6	2.9	490	16	US-10-027-632-38934	Sequence 83646, A
34	41.6	2.9	490	16	US-10-027-632-83646	Sequence 126910,
35	41.6	2.9	800	13	US-10-027-632-126910	Sequence 126910,
36	41.6	2.9	800	16	US-10-027-632-126910	Sequence 45335, A
37	41.6	2.9	885	16	US-10-369-493-45335	Sequence 202, App
38	41.6	2.9	6049	15	US-10-311-455-202	Sequence 2154, App
39	41.6	2.9	7623	15	US-10-311-455-2154	Sequence 28, Appl
40	41.4	2.9	2637	15	US-10-176-306-28	Sequence 32, Appl
41	41	2.9	8323	15	US-10-311-455-32	Sequence 1686, App
42	41	2.9	13326	15	US-10-311-455-1686	Sequence 45, Appl
43	41	2.9	15479	17	US-10-257-166-45	Sequence 1251, App
44	40.8	2.9	7255	15	US-10-311-455-1251	Sequence 1084, App
45	40.8	2.9	10957	15	US-10-311-455-1084	

ALIGNMENTS

RESULT 1
US-09-984-205-5
; Sequence 5, Application US/09984205
; Patent No. US20020137175A1
; GENERAL INFORMATION:
; APPLICANT: Coleman, Timothy A. et al.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: EP0505D1
; CURRENT APPLICATION NUMBER: US/09/984,205
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: PCT/US00/05325
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 03/516,143
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1429
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(590)
US-09-984-205-5

Query Match 100.0%; Score 1429; DB 9; Length 1429;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGGCCTTCCCAAGAAGAACTTCAGGCTCTTGCTGCTGCAACCATCAGCCCAATGACT	60
Db	1	ATGGCCTTCCCAAGAAGAACTTCAGGCTCTTGCTGCTGCAACCATCAGCCCAATGACT	60
Qy	61	GGAATGGAGAAATCACTTTTCAGTAATGGTCAGTATGTTGTAATTCCTTGCAAGAA	120

Db 61 GAGAAATCGAGAAATCAACTTTTCAGTAATTTGGTCGATATGGATTATCTTGTAAGAA 120
 Qy 121 CAGGAGCTGAAGAACATTTTGTGAATGGCAACAACAGAGAGAGCCCTGTCCCTGAGGCTC 180
 Db 121 CAGGAGCTGAAGAACATTTTGTGAATGGCAACAACAGAGAGAGCCCTGTCCCTGAGGCTC 180
 Qy 181 TCAGAGCGTGCAGAGTTCAGAGAGTGGTGTGACAAAGGGAAGCAAGCTGGATCAG 240
 Db 181 TCAGAGCGTGCAGAGTTCAGAGAGTGGTGTGACAAAGGGAAGCAAGCTGGATCAG 240
 Qy 241 GTGATATTCAGCTGAGAGCACTCAGCTTGAAGAGTTCACAGAGTTCAGAGTTCAGAGTTC 300
 Db 241 GTGATATTCAGCTGAGAGCACTCAGCTTGAAGAGTTCACAGAGTTCAGAGTTCAGAGTTC 300
 Qy 301 GCAGAAATAGGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 360
 Db 301 GCAGAAATAGGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 360
 Qy 361 AAGATATTCAGATTAATTTCTTAAGGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 420
 Db 361 AAGATATTCAGATTAATTTCTTAAGGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 420
 Qy 421 TATTACTATCACTTCCTGCTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 480
 Db 421 TATTACTATCACTTCCTGCTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 480
 Qy 481 GGGATTCGGAATAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 540
 Db 481 GGGATTCGGAATAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 540
 Qy 541 GACTTCGGGCAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 Db 541 GACTTCGGGCAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 Qy 601 GATGAGCAACTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 Db 601 GATGAGCAACTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 Qy 661 TCCAGAGATTTATCAACTTTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 720
 Db 661 TCCAGAGATTTATCAACTTTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 720
 Qy 721 TCATGATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 Db 721 TCATGATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 Qy 781 CCAGGAGTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 Db 781 CCAGGAGTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 Qy 841 CTGATTTAAGGATGGAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Db 841 CTGATTTAAGGATGGAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Qy 901 TTGACCTTTGAGACATATCTACCTTAATGATGATGATGATGATGATGATGATGATGATGAT 960
 Db 901 TTGACCTTTGAGACATATCTACCTTAATGATGATGATGATGATGATGATGATGATGATGAT 960
 Qy 961 GAACCTTGAATTAACCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 Db 961 GAACCTTGAATTAACCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 Qy 1021 TGAGCCTTAAAGTTCATTTTGTGAAGGAGGCAAAACTCTAGAGTTCAGAGTTCAGAGTTC 1080
 Db 1021 TGAGCCTTAAAGTTCATTTTGTGAAGGAGGCAAAACTCTAGAGTTCAGAGTTCAGAGTTC 1080
 Qy 1081 TCATTCATTTACAGATTTTGTGGAATTTTGTGGAATTTTGTGGAATTTTGTGGAATTTTGT 1140
 Db 1081 TCATTCATTTACAGATTTTGTGGAATTTTGTGGAATTTTGTGGAATTTTGTGGAATTTTGT 1140
 Qy 1141 AAGAGGAAATTTGATTAATTCATCTCTCTTTAGGAGCTCTCATTTATCTCGGTC 1200
 Db 1141 AAGAGGAAATTTGATTAATTCATCTCTCTTTAGGAGCTCTCATTTATCTCGGTC 1200

Qy 1201 TCTGGTTCCTAATCCTAATTTTAAAGTTGTCTAATTTTAAACCACTATAATATCTTTCAT 1260
 Db 1201 TCTGGTTCCTAATCCTAATTTTAAAGTTGTCTAATTTTAAACCACTATAATATCTTTCAT 1260
 Qy 1261 TTTAATAAATATTTCAATTTTGAATCTAGCAAACTCTGAGCTACTGCAATTTAGGAGGAC 1320
 Db 1261 TTTAATAAATATTTCAATTTTGAATCTAGCAAACTCTGAGCTACTGCAATTTAGGAGGAC 1320
 Qy 1321 TTTAATAAATATTTCAATTTTGAATCTAGCAAACTCTGAGCTACTGCAATTTAGGAGGAC 1380
 Db 1321 TTTAATAAATATTTCAATTTTGAATCTAGCAAACTCTGAGCTACTGCAATTTAGGAGGAC 1380
 Qy 1381 GCTGCTCAGTCTAATCTAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGT 1429
 Db 1381 GCTGCTCAGTCTAATCTAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGT 1429

RESULT 2
 US-09-930-440B-1
 ; Sequence 1, Application US/09930440B
 ; Patent No. US20020142386A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Betenbaugh et al.
 ; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
 ; FILE REFERENCE: PF503P2
 ; CURRENT APPLICATION NUMBER: US/09/930,440B
 ; CURRENT FILING DATE: 2001-08-16
 ; PRIOR APPLICATION NUMBER: 60/227,579
 ; PRIOR FILING DATE: 2000-08-25
 ; PRIOR APPLICATION NUMBER: 09/516,793
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: 60/169,624
 ; PRIOR FILING DATE: 1999-12-08
 ; PRIOR APPLICATION NUMBER: 60/122,582
 ; PRIOR FILING DATE: 1999-03-02
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 1429
 ; TYPE: DNA
 ; ORGANISM: Hmo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(693)
 US-09-930-440B-1

Query Match 100.0%; Score 1429; DB 9; Length 1429;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGSCCTTCCAAAGAGAACTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGT 60
 Db 1 ATGSCCTTCCAAAGAGAACTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGT 60
 Qy 61 GAGATGAGAGAACTCAACTTTTCAGTAAATTCGTTCAGTATGATGATGATGATGATGATGAT 120
 Db 61 GAGATGAGAGAACTCAACTTTTCAGTAAATTCGTTCAGTATGATGATGATGATGATGATGAT 120
 Qy 121 CAGGAGTGAAGAACATTTTGTGAATGGCAACAACAGAGAGGCTTCCTCTGAGGCTC 180
 Db 121 CAGGAGTGAAGAACATTTTGTGAATGGCAACAACAGAGAGGCTTCCTCTGAGGCTC 180
 Qy 181 TCAGAGCTCCCGAGTTCAGAGAGTTCAGAGAGTTCAGAGAGTTCAGAGAGTTCAGAGAGTTCAG 240
 Db 181 TCAGAGCTCCCGAGTTCAGAGAGTTCAGAGAGTTCAGAGAGTTCAGAGAGTTCAGAGAGTTCAG 240
 Qy 241 GTGATATTCAGCTAGGAGCACTGAGCTTGAAGAGTTCAGAGAGTTCAGAGAGTTCAGAGAGTTCAG 300
 Db 241 GTGATATTCAGCTAGGAGCACTGAGCTTGAAGAGTTCAGAGAGTTCAGAGAGTTCAGAGAGTTCAG 300
 Qy 301 GCAGAAATAGGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGT 360

Db 301 GCAGAAATAGGAGCTGATGGGATCGCTGTCATGTGACCGTTCCTCTCAAGCCATGGACC 360
QY 361 AAAGATATCTCTGATTAATTTCTAAAGAAAGTGGTGTCTGCCGCCCTTCCCTTGCATTT 420
Db 361 AAAGATATCTCTGATTAATTTCTAAAGAAAGTGGTGTCTGCCGCCCTTCCCTTGCATTT 420
QY 421 TATTACTATCAGATTCCTGCTGACAGGGGTAAAGATTCGTCTGAGAGTGTGGAT 480
Db 421 TATTACTATCAGATTCCTGCTGACAGGGGTAAAGATTCGTCTGAGAGTGTGGAT 480
QY 481 GGGATTCCTGATAAGATCCCACTTCCAAAGGCTGAATTCAGTGTATACAGATCTCTTA 540
Db 481 GGGATTCCTGATAAGATCCCACTTCCAAAGGCTGAATTCAGTGTATACAGATCTCTTA 540
QY 541 GACTTCGGGCAATGTGTGATCAGAAATCGCCAGCAACAGTTCCTTTTGGGGTG 600
Db 541 GACTTCGGGCAATGTGTGATCAGAAATCGCCAGCAACAGTTCCTTTTGGGGTG 600
QY 601 GATGAGCAACTCTTGAGTCTCTGCTGATGGGCAACTGGAGCAGTGGGAGTGTGTA 660
Db 601 GATGAGCAACTCTTGAGTCTCTGCTGATGGGCAACTGGAGCAGTGGGAGTGTGTA 660
QY 661 TCCAGAGATTTATCAACTTTGTTGTCAAACTAGGTTTGGAGTGTCAAGCCAAAGCCA 720
Db 661 TCCAGAGATTTATCAACTTTGTTGTCAAACTAGGTTTGGAGTGTCAAGCCAAAGCCA 720
QY 721 TCAGACTCTGCTCTGCTGGATTCGATGGGCCACCCGGCTTCCACTGCGAAGCCCT 780
Db 721 TCAGACTCTGCTCTGCTGGATTCGATGGGCCACCCGGCTTCCACTGCGAAGCCCT 780
QY 781 CCAGGGATTTACTGATGCTGAGCTTAACTGAAGAGCCTGGATTTCCCTTCTTTCA 840
Db 781 CCAGGGATTTACTGATGCTGAGCTTAACTGAAGAGCCTGGATTTCCCTTCTTTCA 840
QY 841 CTGATTTAAAGATGAAACTTGGAGCTGGTAGCTAGTGCCTCTCTATCAATCAGGCT 900
Db 841 CTGATTTAAAGATGAAACTTGGAGCTGGTAGCTAGTGCCTCTCTATCAATCAGGCT 900
QY 901 TTGACCTTGAGACATCTACCTTAATAGTGTGATTTTCTCAGGGATTTTAGAT 960
Db 901 TTGACCTTGAGACATCTACCTTAATAGTGTGATTTTCTCAGGGATTTTAGAT 960
QY 961 GAACTTGAACTTAACTCTCTAGCAATGAAATCTCACAATAGCATTTAGSACCTTTTG 1020
Db 961 GAACTTGAACTTAACTCTCTAGCAATGAAATCTCACAATAGCATTTAGSACCTTTTG 1020
QY 1021 TGAGCCTTTAAAGCTTATTTTGTGAAGGGGCAAACTCTAGGAGTCAACACTCTAG 1080
Db 1021 TGAGCCTTTAAAGCTTATTTTGTGAAGGGGCAAACTCTAGGAGTCAACACTCTAG 1080
QY 1081 TCATTCATTTACAGATTTTGTGGAGAAATTTCTGTTATATAGGATGAATGGAATC 1140
Db 1081 TCATTCATTTACAGATTTTGTGGAGAAATTTCTGTTATATAGGATGAATGGAATC 1140
QY 1141 AAGAGAAATTTGATTAATTAATTCCTATCTGTTAGGAGCTCTCATATCTCGGTC 1200
Db 1141 AAGAGAAATTTGATTAATTAATTCCTATCTGTTAGGAGCTCTCATATCTCGGTC 1200
QY 1201 TCTGGTTCCTAATCTTATTTAAAGTGTCTAATTTTAAACCACTATAATATGCTTCAT 1260
Db 1201 TCTGGTTCCTAATCTTATTTAAAGTGTCTAATTTTAAACCACTATAATATGCTTCAT 1260
QY 1261 TTTAATAATATTCATTTGGATCTAGGAAACTCTGAGCTACTGATTTAGGAGGAC 1320
Db 1261 TTTAATAATATTCATTTGGATCTAGGAAACTCTGAGCTACTGATTTAGGAGGAC 1320
QY 1321 TTTAATACCAACTGTAACTGTCTCACTGTATACAACTCAAAATACACCACTCATTT 1380
Db 1321 TTTAATACCAACTGTAACTGTCTCACTGTATACAACTCAAAATACACCACTCATTT 1380
QY 1381 GGTCTGCTCAGTCTAATCTAGAAATGATGCTTTGAAATTCATTTGATG 1423
Db 1381 GGTCTGCTCAGTCTAATCTAGAAATGATGCTTTGAAATTCATTTGATG 1423

RESULT 3
US-10-102-524-1751
; Sequence 1751, Application US/10102524
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; THERAPY AND DIAGNOSIS OF KIDNEY CANCER
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1751
; LENGTH: 1552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-1751

Query Match 92.9%; Score 1328; DB 15; Length 1552;
Best Local Similarity 94.4%; Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 0; Indels 85; Gaps 1;

QY 1 ATGGCCCTTCCCAAGAGAAACTTTCAGGGTCTTGTGGCTGCAACCATCAGCCCAATGACT 60
Db 45 ATGGCCCTTCCCAAGAGAAACTTTCAGGGTCTTGTGGCTGCAACCATCAGCCCAATGACT 104
QY 61 GAGAAATGAGAAATCAACTTTTCAGTAAATTTGTCAGTATGTGATTTCTTGTGAAGAA 120
Db 105 GAGAAATGAGAAATCAACTTTTCAGTAAATTTGTCAGTATGTGATTTCTTGTGAAGAA 164
QY 121 CAGGGATGCAAGCAATTTTGTGAATGCAACAGGAGAGGCTGTCCCTGAGGCTC 180
Db 165 CAGGGATGCAAGCAATTTTGTGAATGCAACAGGAGAGGCTGTCCCTGAGGCTC 224
QY 181 TCAGAGCTCGCCAGGTTTCAGAGAGTGGTGCACAAAAGGAGGAGCAAGCTGGATCAG 240
Db 225 TCAGAGCTCGCCAGGTTTCAGAGAGTGGTGCACAAAAGGAGGAGCAAGCTGGATCAG 284
QY 241 GTGATAATTCAGTAGAGCACTGAGCTTGAAGAGTGCAGGAACTGGCCCAACATGCA 300
Db 285 GTGATAATTCAGTAGAGCACTGAGCTTGAAGAGTGCAGGAACTGGCCCAACATGCA 344
QY 301 GCAGAAATAGGAGCTGATGGCATCGCTGTCATTCACCCCTTCTTCTCAAGCCATGGACC 360
Db 345 GCAGAAATAGGAGCTGATGGCATCGCTGTCATTCACCCCTTCTTCTCAAGCCATGGACC 404
QY 361 AAAGATATCTGATTAATTTCTTAAAGAAAGTGGTGTCTGCCGCCCTTCCCTTSCCATTT 420
Db 405 AAAGATATCTGATTAATTTCTTAAAGAAAGTGGTGTCTGCCGCCCTTCCCTTSCCATTT 464
QY 421 TATTACTATCAGATTCCTGCTTGAAGAGTAAAGATTCGTCTGAGAGTGTGGAT 480
Db 465 TATTACTATCAGATTCCTGCTTGAAGAGTAAAGATTCGTCTGAGAGTGTGGAT 524
QY 481 GGGATTCCTGATAAGATCCCACTTCCAAAGGCTGAATTCAGTGTATACAGATCTCTTA 540
Db 525 GGGATTCCTGATAAGATCCCACTTCCAAAGGCTGAATTCAGTGTATACAGATCTCTTA 584
QY 541 GACTTCGGGCAATGTGTGATCAGAAATCGCCAGCAACAGTTCCTTTTGGGGTG 600
Db 585 GACTTCGGGCAATGTGTGATCAGAAATCGCCAGCAACAGTTCCTTTTGGGGTG 644
QY 601 GATGAGCAACTCTTGAGTGTCTCTGATGATGGGAGCAACTGGAGCACTGGG 650
Db 645 GATGAGCAACTCTTGAGTGTCTCTGATGATGGGAGCAACTGGAGCACTGGGAGCACTGAT 704

```
Qy 651 ----- 650
Db 705 AACTACCTGGGAAAAAGACAAACCCAGATGTTGGAGGCTTTTSAACAAAGGACTTCCT 764
Qy 651 -----CAGTTTGTATCCAGAGATTATCAACTTGTGTCAAACTAGGT 695
Db 765 TTAGCCCTGAACTATCAGTTTGTATCCAGAGATTATCAACTTGTGTCAAACTAGGT 824
Qy 696 TTGGAGTGTACAGACCAAGCCATCATGATCTCTGCTCTGGATTCGAATCCAAATGGCCCA 755
Db 825 TTGGAGTGTACAGACCAAGCCATCATGATCTCTGCTCTGGATTCGAATCCAAATGGCCCA 884
Qy 756 CCCGGCTTCCACTGCAGAAAGCCCTCCAGGGAGTTTACTGATAGTGCCTGAAGCTAAATCTG 815
Db 885 CCCGGCTTCCACTGCAGAAAGCCCTCCAGGGAGTTTACTGATAGTGCCTGAAGCTAAATCTG 944
Qy 816 AAGAGCTTGATTTCTCTTTTCACTGATTTAAAGGATGAAATCTTGAAGCTGGTAGC 875
Db 945 AAGAGCTTGATTTCTCTTTTCACTGATTTAAAGGATGAAATCTTGAAGCTGGTAGC 1004
Qy 876 TAGTGCCTCTCTNCAATCAGGTTTGCACCTTGGAGACNTAATCTACCTTAATAGTAGC 935
Db 1005 TAGTGCCTCTCTNCAATCAGGTTTGCACCTTGGAGACNTAATCTACCTTAATAGTAGC 1064
Qy 936 ATTTTCTTCCAGGATTTTAGATGAATCTTGAATTAATCTCTTGAAGCTGAAATCTC 995
Db 1065 ATTTTCTTCCAGGATTTTAGATGAATCTTGAATTAATCTCTTGAAGCTGAAATCTC 1124
Qy 996 ACAATAAGCAATGAGGTACCTTTTGTAGCCCTTAAAGAGTCTTATTTTGTGAAGGGCAA 1055
Db 1125 ACAATAAGCAATGAGGTACCTTTTGTAGCCCTTAAAGAGTCTTATTTTGTGAAGGGCAA 1184
Qy 1056 AAACCTAGGAGTCACAACTCTCAGTCACTTCAATTCACAGATTTTGTGAGAGAAATTT 1115
Db 1185 AAACCTAGGAGTCACAACTCTCAGTCACTTCAATTCACAGATTTTGTGAGAGAAATTT 1244
Qy 1116 CTGTTTATATGATGAAATGGAATCAAGAGAAATTTGAATTAATTAATTCATCTCTC 1175
Db 1245 CTGTTTATATGATGAAATGGAATCAAGAGAAATTTGAATTAATTAATTCATCTCTC 1304
Qy 1176 TTTAGGAGCTCTCAATTAATCTCGGCTCTGTTTCTTAATTCCTTAATTTAAAGTTTCTAAT 1235
Db 1305 TTTAGGAGCTCTCAATTAATCTCGGCTCTGTTTCTTAATTCCTTAATTTAAAGTTTCTAAT 1364
Qy 1236 TTAACCACTAATATATGCTCTCAATTTAATAATTAATTCATTTGGAATCTAGAAATCT 1295
Db 1365 TTAACCACTAATATATGCTCTCAATTTAATAATTAATTCATTTGGAATCTAGAAATCT 1424
Qy 1296 TGAGTACTGCAATTTAGGAGGCACTTTAATACCAATCTGTAACATGCTCAACTGTATA 1355
Db 1425 TGAGTACTGCAATTTAGGAGGCACTTTAATACCAATCTGTAACATGCTCAACTGTATA 1484
Qy 1356 CAATCAAAATACACAGCTCATTTGGTGTCTAGTCTAATCTTAGAATGGAATGCTTTTG 1415
Db 1485 CAATCAAAATACACAGCTCATTTGGTGTCTAGTCTAATCTTAGAATGGAATGCTTTTG 1544
Qy 1416 AATTCATT 1423
Db 1545 AATTCATT 1552
```

RESULT 4
US-09-833-381-1100/c
; Sequence 1100, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29

; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1190
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-1100

Query Match 55.4%; Score 792; DB 9; Length 1923;
Best Local Similarity 91.2%; Pred. No. 2.1e-217;
Matches 899; Conservative 0; Mismatches 0; Indels 87; Gaps 2;

```
Qy 531 AGATCTCTTAGACTTCGGCAATGTGTGATCAGCAATGCCAGCAACAGTTTGTCTTCTTCT 590
Db 1923 AGATCTCTTAGACTTCGGCAATGTGTGATCAGCAATGCCAGCAACAGTTTGTCTTCTTCT 1864
Qy 591 TTTTGGGCTGGATGAGCAACTGTTGAGTGCTCTGGTGATGGGAGCAACTGGAGCAGTGG 650
Db 1863 TTTTGGGCTGGATGAGCAACTGTTGAGTGCTCTGGTGATGGGAGCAACTGGAGCAGTGG 1804
Qy 651 ----- 650
Db 1803 CAGTACCTATAACTACCTGGGAAAAACACAAACCAGATGTTGGGGGCTTTTGAACAAAA 1744
Qy 651 -----CAGTTTGTATCCAGAGATTATCAACTTTGTG 684
Db 1743 GGGACTTCTCTTTAGCCCTGAACTATCAGTTTGTATCCAGAGATTATCAACTTTGTG 1684
Qy 685 TCAAACTAGGTTTGGAGTGTACAGACCAAGCCATCATGCTCTGCTCTGGGATTC 744
Db 1683 TCAAACTAGGTTTGGAGTGTACAGACCAAGCCATCATGCTCTGCTCTGGGATTC 1624
Qy 745 CAATGGGCGCACCCCGGCTTCCACTGCAGAAAGCCTCCAGGAGTTTACTGATGCTG 804
Db 1623 CAATGGGCGCACCCCGGCTTCCACTGCAGAAAGCCTCCAGGAGTTTACTGATGCTG 1564
Qy 805 AAGCTAAACTCAAGAGCCTGGATTTCCCTTTCTTCACTGATTTAAAGGATGAAACTTGG 864
Db 1563 AAGCTAAACTCAAGAGCCTGGATTTCCCTTTCTTCACTGATTTAAAGGATGAAACTTGG 1504
Qy 865 AAGCTGGTAGTGGCTCTCTATCAAAATCAGGTTTGGACCTTTGAGACATAATCTACC 924
Db 1503 AAGCTGGTAGTGGCTCTCTATCAAAATCAGGTTTGGACCTTTGAGACATAATCTACC 1444
Qy 925 TTAATAGTGCATTTTCTCAGGAGTTTGTAGATGAATTAATTAATCTCTCTCTCA 984
Db 1443 TTAATAGTGCATTTTCTCAGGAGTTTGTAGATGAATTAATTAATCTCTCTCTCA 1384
Qy 985 AATGAATCTCACAATAGCAATGAGGTACCTTTTGTGAGCCTTTAAAGTCTTATTTTG 1044
Db 1383 AATGAATCTCACAATAGCAATGAGGTACCTTTTGTGAGCCTTTAAAGTCTTATTTTG 1324
Qy 1045 TGAAGGGCAAAAACTCTAGGAGTCACAACTCTCAGTCATTCATTTCTCAGATTTTGTG 1104
Db 1323 TGAAGGGCAAAAACTCTAGGAGTCACAACTCTCAGTCATTCATTTCTCAGATTTTGTG 1264
Qy 1105 TGGAGAAATTTCTCTTATATGATGAAATCGAATCAAGAGAAATTTGTAATGATTAA 1164
Db 1263 TGGAGAAATTTCTCTTATATGATGAAATCGAATCAAGAGAAATTTGTAATGATTAA 1204
Qy 1165 TTCCATCTGCTTTTAGGAGCTCTCATTTATCTCGGCTCTGCTCTCTTAATCTTATTTAAA 1224
Db 1203 TTCCATCTGCTTTTAGGAGCTCTCATTTATCTCGGCTCTGCTCTCTTAATCTTATTTAAA 1144
Qy 1225 GTTCTCTAATTTTAAACCACTATATATGCTCTCATTTTAAATAATTCATTTTGAATC 1284
Db 1143 GTTCTCTAATTTTAAACCACTATATATGCTCTCATTTTAAATAATTCATTTTGAATC 1084
Qy 1285 TAGGAATCTCTGAGCTACTGCAATTTAGGAGGCACTTTAATACCACTGTACATCTC 1344
Db 1083 TAGGAATCTCTGAGCTACTGCAATTTAGGAGGCACTTTAATACCACTGTACATCTC 1024
Qy 1345 TCAACTGNTACAACTCAAAATACACCAAGCTCAATTTGGCTGCTCAGTCTAATCTTAGAAT 1404
```

Db 1023 TCAACTGTATCAACTCAAAATACACAGCTCATTTGGCTGCTCAGTCTAACTCTAGAAAT 964
 QY 1405 -GGATGCTTTGAATTCATTTCCGATG 1429
 Db 963 GGGAGCTTTTGAATTCATTTCCGATG 938

RESULT 5
 US-09-833-381-1099
 ; Sequence 1099, Application US/09833381
 ; Patent No. US20020132090A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Robison, Keith E.
 ; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
 ; FILE REFERENCE: 5800-119
 ; CURRENT APPLICATION NUMBER: US/09/833,381
 ; CURRENT FILING DATE: 2001-04-11
 ; PRIOR APPLICATION NUMBER: 09/516,448
 ; PRIOR FILING DATE: 2000-02-29
 ; NUMBER OF SEQ ID NOS: 2050
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1099
 ; LENGTH: 1127
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-833-381-1099

Query Match 45.8%; Score 654; DB 9; Length 1127;
 Best Local Similarity 100.0%; Pred. No. 1.1e-177; Mismatches 0; Indels 0; Gaps 0;
 Matches 654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCCAAGAGAACTTCAGGCTCTTGGGTGCAACCATCACGCGCAATGACT 60
 Db 199 ATGGCTTCCCAAGAGAACTTCAGGCTCTTGGGTGCAACCATCACGCGCAATGACT 258

QY 61 GAGATGAGAAATCAACTTTTCATATATGTCAGTATGTCATATCTTGTGAAGAA 120
 Db 259 GAGATGAGAAATCAACTTTTCATATATGTCAGTATGTCATATCTTGTGAAGAA 318

QY 121 CAGGAGTGAAGAACTTTTGTGATGCGCAACAGAGAGAGGCTGTCTCCCTGAGCGTC 180
 Db 319 CAGGAGTGAAGAACTTTTGTGATGCGCAACAGAGAGAGGCTGTCTCCCTGAGCGTC 378

QY 181 TCAGAGCTGCGCAGGTTGCAGAGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAG 240
 Db 379 TCAGAGCTGCGCAGGTTGCAGAGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAG 438

QY 241 GTGATTAATTCAGTAGAGCACTGAGCTTGAAGAGATGACAGAACTGGCCCAACATGA 300
 Db 439 GTGATTAATTCAGTAGAGCACTGAGCTTGAAGAGATGACAGAACTGGCCCAACATGA 498

QY 301 CGAGAAATAGAGCTGATGGCATCCTGTCATTGCAACGCTTCTCTCAAGCCATGGACC 360
 Db 499 CGAGAAATAGAGCTGATGGCATCCTGTCATTGCAACGCTTCTCTCAAGCCATGGACC 558

QY 361 AAGATATCTCTGATTAATTTCTAAGGAAGTGGCTGCTGCGCCCTGCGCTGCCATTT 420
 Db 559 AAGATATCTCTGATTAATTTCTAAGGAAGTGGCTGCTGCGCCCTGCGCTGCCATTT 618

QY 421 TATTACTATCAATTCCTGCTTGAAGAGTGAAGAGTTCGCTGAGAGAGTTTGGAT 480
 Db 619 TATTACTATCAATTCCTGCTTGAAGAGTGAAGAGTTCGCTGAGAGAGTTTGGAT 678

QY 481 GGGATTCGGATAGATCCCACTTCCAAAGGCTGAAATTCAGATGATACAGATCTCTTA 540
 Db 679 GGGATTCGGATAGATCCCACTTCCAAAGGCTGAAATTCAGATGATACAGATCTCTTA 738

QY 541 GACTTCGGGCAATGTGTTGATCAGAAATCGCCAGCAACAGTTTGTCTTTTGGGGTG 600
 Db 739 GACTTCGGGCAATGTGTTGATCAGAAATCGCCAGCAACAGTTTGTCTTTTGGGGTG 798

QY 601 GATGAGCAACTGTTGAGTGTCTGCTGATGGAGCAACTGAGAGCAGTGGGCAGT 654

Db 799 GATGAGCAACTGTTGAGTGTCTGCTGATGGAGCAACTGAGAGCAGTGGGCAGT 852

RESULT 6
 US-10-102-524-220/c
 ; Sequence 220, Application US/10102524
 ; Publication No. US20030109434A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Xanion, Jane
 ; APPLICANT: Gaiger, Alexander
 ; APPLICANT: Gordon, Brian
 ; APPLICANT: Harlocker, Susan L.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 ; THERAPY AND DIAGNOSIS OF KIDNEY CANCER
 ; FILE REFERENCE: 210121.572
 ; CURRENT APPLICATION NUMBER: US/10/102,524
 ; CURRENT FILING DATE: 2002-03-19
 ; NUMBER OF SEQ ID NOS: 1863
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 220
 ; LENGTH: 374
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-102-524-220

Query Match 26.2%; Score 374; DB 15; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.9e-97; Mismatches 0; Indels 0; Gaps 0;
 Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 848 AAGAGTGAAGAACTTGGAGCTGGTAGCTAGTGCCTCTCTATCAATCAGGTTGCACC 907
 Db 374 AAGAGTGAAGAACTTGGAGCTGGTAGCTAGTGCCTCTCTATCAATCAGGTTGCACC 315

QY 908 TTGACACATATCTACCTTAATAGTGCATTTTTCAGGAAATTTAGTGAATCTG 967
 Db 314 TTGACACATATCTACCTTAATAGTGCATTTTTCAGGAAATTTAGTGAATCTG 255

QY 968 AATAAATCTCTAGCAATGAATGAATCTCAATATAGCATTTGAGTACCTTTGTAGCCCT 1027
 Db 254 AATAAATCTCTAGCAATGAATGAATCTCAATATAGCATTTGAGTACCTTTGTAGCCCT 195

QY 1028 TAAAGATCTTATTTTGAAGGCGCAAAACTCTAGAGTGCACAACTCTCAGTCTCA 1087
 Db 194 TAAAGATCTTATTTTGAAGGCGCAAAACTCTAGAGTGCACAACTCTCAGTCTCA 135

QY 1088 TTTCACAGATTTTGTGGAGAAATTTCTGTTATATGATGAATGAATCAAGAGGA 1147
 Db 134 TTTCACAGATTTTGTGGAGAAATTTCTGTTATATGATGAATGAATCAAGAGGA 75

QY 1148 AAATGTTAATGATTAATTCATCTGCTTTAGGAGCTCTCATTTCTGGTCTCTGGTT 1207
 Db 74 AAATGTTAATGATTAATTCATCTGCTTTAGGAGCTCTCATTTCTGGTCTCTGGTT 15

QY 1208 CCTAATCTTATTTT 1221
 Db 14 CCTAATCTTATTTT 1

RESULT 7
 US-10-085-783A-21707
 ; Sequence 21707, Application US/10085783A
 ; Publication No. US20040037841A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ChondroGene Inc.
 ; APPLICANT: Liew, C.C.
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
 ; FILE REFERENCE: 4231/2002
 ; CURRENT APPLICATION NUMBER: US/10/085,783A
 ; CURRENT FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 60/305,340
 ; PRIOR FILING DATE: 2001-07-13


```

; LOCATION: (213)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (214)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (228)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (231)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (267)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (292)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (300)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-1834

Query Match          10.4%; Score 149.2; DB 9; Length 309;
Best Local Similarity 95.6%; Pred. No. 3e-32;
Matches 195; Conservative 0; Mismatches 5; Indels 4; Gaps 4;

QY 322 ATCGCTGTCATTGCACGGTTCCTCTCAAGCCATGACCAAGATATCTGATTATTC 381
DB 9 ATCGCTGTCATTGCACGGTTCCTCTCAAGCCATGACCAAGATATCTGATTATTC 68

QY 382 CTAAGGAGTGGCTGCTGCCGCCCTGCTGCCCTGCTGCCCTGCTGCCCTGCTGCC 441
DB 69 CTAAGGAGTGGCTGCTGCCGCCCTGCTGCCCTGCTGCCCTGCTGCCCTGCTGCC 128

QY 442 TTG-ACAGGGTAAAGATTTCGTG-CTGAGGAGTTGTTGGATGG-ATTCTGGATAGATC 498
DB 129 TTGANAAGGGTAAAGATTTCGTG-CTGAGGAGTTGTTGGATGG-ATTCTGGATAGATC 188

QY 499 CCCACC-TTCCAAAGGGCTGAAATT 521
DB 189 CCCACC-TTCCAAAGGGCTGAAATT 212

RESULT 10
US-09-918-995-8879
; Sequence 8879, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8879
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(472)
; OTHER INFORMATION: n = A,T,C or G
; US-09-918-995-8879

Query Match          9.0%; Score 128.2; DB 10; Length 472;
Best Local Similarity 97.7%; Pred. No. 4.8e-26;
Matches 130; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1297 GAGTACTGTCATTAGGCGAGGCACTTTAATACCAACTGTACATGTCTCAACTGTATAC 1356

US-09-930-440b-1.rnpb

Db 61 GAATTCCTGCATTAGGCGAGGCACTTTAATACCAACTGTACATGTCTCAACTGTATAC 120
QY 1357 AACTCAAAATACACACAGCTCATTGGCTGCTAGTCTTAACCTAGATGATGCTTTTGA 1416
DB 121 AACTCAAAATACACACAGCTCATTGGCTGCTAGTCTTAACCTAGATGATGCTTTTGA 180
QY 1417 ATTCATTTCGATG 1429
DB 181 ATTCATTTCGATG 193

RESULT 11
US-10-085-117-223
; Sequence 223, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 223
; LENGTH: 177249
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(177249)
; OTHER INFORMATION: n = any nucleotide
; US-10-085-117-223

Query Match          3.2%; Score 46.4; DB 16; Length 177249;
Best Local Similarity 45.6%; Pred. No. 1.3;
Matches 164; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

QY 952 ATTTCAGTGAATCAATAAACTCTCCCTAGCAAAATGAAATCTCACAATTAAGCATTCAGG 1011
DB 71301 ACTTTACCTGCGAGTTTACTGGCTACTGTTCAAGAGAGAGTTCTTTGCTTATGACTGGCT 71360

QY 1012 TACCTTTTGTGAGCCTTAAAAAGTCTTATTTTGTGAAGGGGCAAAAACCTTAGGAGTCA 1071
DB 71361 TACCTTAGATAATTCCTTAAACTATATATATATATATATATATATACACACACATACATAC 71420

QY 1072 AACTCTCAGTCAATTCATTTTCAGAGATTTTGTGGAGAAATTCGTGTTTATATGATGA 1131
DB 71421 ACACACACATATACATACACACACATATGATGATATATATATATATATATATATATGATGT 71480

QY 1132 AATGGAATCAGAGGAAATTCGTAATGTAATTCATTCCTCTCTCTTTAGGAGCTCTCATT 1191
DB 71481 AATACACACATATATATATATATATATATATATATATATATATATATATATATATATAT 71540

QY 1192 ATCTCGGCTCTGTTTCCTAATCCCTAATTTTAAAGTTGTCTAATTTTAAACCCATATA 1251
DB 71541 ATGTTATTAACAATATGATTTTAAATATAAAATCTTAATATAATTAATATA 71600

QY 1252 TGTCTTCATTTTAAATAATTCATTTCGGAATCTAGGAAACCTCGAGCTACTGCATTTA 1311
DB 71601 CATATACACATATATATATATATATATATATATATATATATATATATATATATATAT 71660

RESULT 12
US-10-311-455-998
; Sequence 998, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

```

```
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 998
; LENGTH: 6117
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: 4521
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-998

Query Match      3.2%; Score 45.4; DB 15; Length 6117;
Best Local Similarity 49.5%; Pred. No. 0.24; 146; Indels 1; Gaps 1;
Matches 144; Conservative 0; Mismatches 0;

QY 992 TCTCAGCAATAGCATTGAGTACCTTTTGTGAGCCTTAAAGCTTATTTTGTGAAGGG 1051
Db 316 TATTAGTATATGATTTATTTAGTGTAGTTGGGAAGATATATATTTGATTTTAGGTA 375
QY 1052 GCAAAACTAGGAGTCACAACTCTCAGTCATTCATTTCACAGATTTTGTGGAGAA 1111
Db 376 GAAATATCTTTTCTATTATTTGTTTATTTTATTTTGTGAAATTTATCGTGAAGTT 435
QY 1112 ATTCTCTTTATATGATGAATCGAATCAAGAGAAATTTGATTAATTCATTCCTC 1171
Db 436 AATATATTATTTTGGAGAGAGATTTATAGAAATTTATGTTGTTATTAATTTATGTTA 495
QY 1172 TGCTTTAGGAGCTCTCATTAATCTCGGTCTCTGTTCTATCTATCTATTTAAAGTTGCT 1231
Db 496 GAAGTGCTGTTGTTGTTATT-GTAGGATTTATTAATAAATTCGATTTTATGTTT 554
QY 1232 AATTTAAACCATATATATATGCTTCACTTTTAAATATATTCATTCATTCGAA 1282
Db 555 TTATAGCTATATTTTATGTTGTTTAAATTTAATATGATGATTTGTTGTA 605

RESULT 13
US-10-311-455-1924
; Sequence 1924, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1924
; LENGTH: 5771
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (214, 2796..2797, 4347)
US-10-221-613-33
Query Match      3.1%; Score 44.4; DB 13; Length 6109;
Best Local Similarity 52.1%; Pred. No. 0.47;
Matches 99; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1089 TTCAAGATTTTGTGGAGAAATTTCTGTTATATGATGAATGAATGAATCAAGAGAA 1148
Db 1641 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1700
QY 1149 AATTGTAATTCATTAATTCATCTCTTTAGGAGCTCTCATTAATTCGGTCTCGGTTTC 1208
Db 1701 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1760
QY 1209 CTAATCCTATTAAAGTTGCTGAATTTTAAACCATATAATATGCTCTCTCAATTTAATA 1268
Db 1761 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1820
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1924

Query Match      3.1%; Score 44.8; DB 15; Length 5771;
Best Local Similarity 55.0%; Pred. No. 0.35;
Matches 88; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1079 AGTCATTTCATTCACAGATTTTGTGGAGAAATTTCTGTTATATGGAATCAATCGAA 1138
Db 1803 AGTTTGTGTTTGTAGATATTTTGTGAAAGTAGTTTATATATATTAATTAAGGTTA 1852
QY 1139 TCAGAGGAATTAATTAATTCATTCATCTGCTTTAGGAGCTCTCATTAATTCCTCGG 1198
Db 1863 AAAAGTAGATATTAGTAATTTTAAATGATGATATATGTTGTTATTATTATTATT 1922
QY 1199 TCTCTGTTCTCTAATCTCTATTTTAAAGTTGTCTAATTTTA 1238
Db 1923 TGCGCTTTTATTATTATTATTATTATTATTATTAAAGTGA 1962
```

```
RESULT 14
US-10-221-613-33
; Sequence 33, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019258.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 33
; LENGTH: 6109
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (214, 2796..2797, 4347)
US-10-221-613-33
```

```
Query Match      3.1%; Score 44.4; DB 13; Length 6109;
Best Local Similarity 52.1%; Pred. No. 0.47;
Matches 99; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1089 TTCAAGATTTTGTGGAGAAATTTCTGTTATATGATGAATGAATGAATCAAGAGAA 1148
Db 1641 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1700
QY 1149 AATTGTAATTCATTAATTCATCTCTTTAGGAGCTCTCATTAATTCGGTCTCGGTTTC 1208
Db 1701 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1760
QY 1209 CTAATCCTATTAAAGTTGCTGAATTTTAAACCATATAATATGCTCTCTCAATTTAATA 1268
Db 1761 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1820
```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:55:49 ; Search time 32,425.1 Seconds
(without alignments)
2004.184 Million cell updates/asc

Title: US-09-930-440B-2

Perfect score: 1159

Sequence: 1 MAFPKKKLQGLVAATIPMT.....GATGAGSVFVRDLSTLSLN 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1159	100.0	230	3	AA96099 Human ald
2	1159	100.0	230	3	AA96099 Human ald
3	1159	100.0	230	4	AA96099 Human ald
4	1159	100.0	230	6	AA96099 Human ald
5	1104.5	95.3	284	5	AA96099 Human ald
6	1104.5	95.3	284	5	AA96099 Human ald
7	1104.5	95.3	320	2	AA96099 Human ald
8	1104.5	95.3	320	2	AA96099 Human ald
9	718.5	62.0	301	5	AA96099 Human ald
10	675	58.2	147	4	AA96099 Human ald
11	599	51.7	141	4	AA96099 Human ald
12	274.5	23.7	297	2	AA96099 Human ald
13	274.5	23.7	297	2	AA96099 Human ald
14	272.5	23.5	297	4	AA96099 Human ald
15	254.5	22.0	297	4	AA96099 Human ald
16	252.5	21.8	305	6	AA96099 Human ald
17	252.5	21.8	305	6	AA96099 Human ald
18	247	21.3	305	5	AA96099 Human ald
19	242.5	20.9	293	6	AA96099 Human ald
20	241.5	20.8	341	6	AA96099 Human ald
21	239.5	20.7	302	7	AA96099 Human ald
22	237.5	20.5	291	6	AA96099 Human ald
23	233.5	19.3	289	6	AA96099 Human ald
24	222.5	19.2	301	2	AA96099 Human ald
25	222.5	19.2	301	2	AA96099 Human ald

26	222.5	19.2	301	2	AAW23283
27	222.5	19.2	301	2	AAW06582
28	222.5	19.2	301	2	AAW06582
29	222.5	19.2	301	2	AAW06582
30	222.5	19.2	301	2	AAW06582
31	222.5	19.2	301	2	AAW06582
32	222.5	19.2	301	2	AAW06582
33	222.5	19.2	301	2	AAW06582
34	222.5	19.2	301	2	AAW06582
35	222.5	19.2	301	2	AAW06582
36	222.5	19.2	301	2	AAW06582
37	222.5	19.2	301	2	AAW06582
38	222.5	19.2	301	2	AAW06582
39	222.5	19.2	301	2	AAW06582
40	219	18.9	298	6	ABM38742
41	218.5	18.9	319	5	ABU17664
42	218.5	18.9	1006	4	ABU15195
43	218.5	18.9	1131	4	ABG17695
44	218	18.8	309	6	ABG24897
45	215	18.6	304	5	ABU28508

ALIGNMENTS

RESULT 1

AA96099

ID AA96099 standard; protein; 230 AA.

AC AA96099;

XX

DT 19-DEC-2000 (first entry)

XX

DE Human aldolase.

XX

KW Aldolase; human; sialylation; glycoprotein; plasminogen; transferrin;

KW thyrotropin; Na+,K+-ATPase.

XX

OS Hemo sapiens.

XX

PN WO200052135-A2.

XX

PD 08-SEP-2000.

XX

PF 01-MAR-2000; 2000WO-US005313.

XX

PR 02-MAR-1999; 95US-0122582P.

PR 08-DEC-1999; 95US-0169624P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PA (UYJO) UNIV JOHNS HOPKINS.

XX

PI (UYWY-) UNIV WYOMING.

XX

PT Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;

XX

DR MPI; 2000-572178/53.

XX

DR N-PSDB; AAA50567.

XX

PT Recombinant production of sialylated glycoproteins using cells in which

XX

PT the expression of enzymes, e.g. sialic acid synthetase, involved in the

XX

PT sialylation reaction has been altered.

XX

ES Claim 18; Page 100; 144pp; English.

XX

CC The present sequence is that of human aldolase. The invention provides

CC methods and recombinantly engineered cells for producing glycoproteins

CC having sialylated oligosaccharides. The methods involve altering the

CC expression of enzymes involved in carbohydrate processing. A claimed cell

CC producing sialylated glycoprotein at above endogenous levels expresses at

CC least 1 (preferably human) enzyme selected from GlcNAc-2 epimerase, an

CC enzyme that catalyzes the conversion of UDP-GlcNAc to ManNAc, sialic acid

CC synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid

CC transporter at above endogenous levels. Endogenous N-

CC acetylglucosaminidase activity may be suppressed. A claimed method for
CC manipulating glycoprotein in an insect cell comprises enhancing the
CC expression of 1 of the above enzymes, and a claimed method for producing
CC sialylated glycoproteins involves expressing a heterologous protein
CC (especially plasminogen, transferrin, Na⁺, K⁺-ATPase or thyrotropin) in
CC the insect cell. Yeast, insect, fungal, plant and bacterial host cells
CC can be engineered to produce new forms of sialylated glycoproteins,
CC higher concentrations of sialylated glycoproteins and/or elevated
CC concentrations of donor substrates (e.g. nucleotide sugars) required for
CC sialylation
XX
SQ

Query Match 100.0%; Score 1159; DB 3; Length 230;
Best Local Similarity 100.0%; Pred. No. 8.1e-120;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAFPKKQLGLVAATITPTMTENGESVIGQYDYLKQGVKNIFVNGTIGGLSLSV 60
Db - MAFPKKQLGLVAATITPTMTENGESVIGQYDYLKQGVKNIFVNGTIGGLSLSV 60
QY 61 SERQVAEWTGKOKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPEWT 120
Db 61 SERQVAEWTGKOKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPEWT 120
QY 121 KDILINFLKEVAAAAPALPFYVYHHPALTGKIRABELLDGILDKIPTFQGLKESDIDL 180
Db 121 KDILINFLKEVAAAAPALPFYVYHHPALTGKIRABELLDGILDKIPTFQGLKESDIDL 180
QY 181 DFGQCVQDNRQQQAFALFGVDEQLLSALVMGATGAVGSFVSRDLSTLLSN 230
Db 181 DFGQCVQDNRQQQAFALFGVDEQLLSALVMGATGAVGSFVSRDLSTLLSN 230

RESULT 2
AAY9C353
ID AAY90353 standard; protein; 230 AA.
XX AAY90353;
DT 04-DEC-2000 (first entry)
DE Human glycosylation enzyme clone HDPAX85 protein sequence.
XX Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
KW immunotherapy; cosmetic surgery; metabolism; immune system disorder;
KW haematopoietic cell deficiency; blood coagulation disorder; asthma;
KW afibrinogenemia; blood platelet disorder; thrombocytopaenia; neoplasia;
KW autoimmune disorder; Addison's disease; multiple sclerosis; purpura;
KW allergic encephalomyelitis; allergic reaction; organ rejection;
KW graft-versus-host disease; inflammation; hyperproliferative disorder;
KW sarcoidosis; infection; gene therapy; CMP sialic acid synthetase.
XX
OS Homo sapiens.
XX WO200052136-A2.
XX
PD 08-SEP-2000.
XX
PF 01-MAR-2000; 2000WO-US005325.
XX
PR 02-MAR-1999; 99US-0122409P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Coleman TA;
XX
XW MPI; 2000-572179/53.
DR N-PSDB; AAA37764.
XX
PT New human glycosylation enzymes cytidine 5'-monophosphate sialic acid
PT synthetase, sialic acid synthetase and aldolase and nucleic acids
PT encoding the proteins for treating e.g., immune system disorders,

PT microbial diseases.
XX
PS Claim 12; Page 113-114; 115pp; English.
XX
CC This sequence represents a human glycosylation enzyme clone of the
CC invention, designated HDPAX85. This protein clone is an aldolase. The
CC sequences are useful as reagents for the differential identification of
CC the tissues or cell types present in a biological sample, as
CC immunological probes, for treating a disease or condition resulting from
CC under expression of such polypeptide, for the detection and/or treatment
CC of disorders involving aberrant glycolysis, e.g. cramps, myoglobinuria,
CC and as tumour marker and/or immunotherapy targets. They may also be used
CC to differentiate, proliferate and attract cells leading to the
CC regeneration of tissues, to modulate mammalian metabolism affecting catabolism, anabolism
CC (cosmetic surgery) or mammalian metabolism affecting catabolism, anabolism
CC processing, utilisation and energy storage, to change a mammal's mental
CC state by influencing and as a food additive or preservative. The proteins
CC can be used to assay protein levels in a sample, as a marker or detector
CC of an immune system disorder, to inhibit cytokine activity, and as a
CC vaccine. They may further be used to treat immune system or of
CC haematopoietic cell deficiencies or disorders, blood coagulation
CC disorders (e.g. afibrinogenemia), blood platelet disorders (e.g.
CC thrombocytopaenia) wounds resulting from trauma or surgery, allergic
CC disorders (e.g. Addison's disease, multiple sclerosis, allergic
CC encephalomyelitis), allergic reactions (e.g. asthma), organ rejection,
CC graft-versus-host disease, inflammation, hyperproliferative disorders
CC (e.g. neoplasia, purpura, sarcoidosis), diseases caused by viruses (e.g.
CC hepatitis, meningitis, AIDS), bacteria and fungi (include e.g.
CC tuberculosis, conjunctivitis, sepsis, typhoid, chlamydia, cellulitis),
CC and diseases caused by parasites (e.g. amoebiasis, coccidiosis,
CC leishmaniasis, scabies, malaria, toxoplasmosis)
XX
SQ Sequence 230 AA;

Query Match 100.0%; Score 1159; DB 3; Length 230;
Best Local Similarity 100.0%; Pred. No. 8.1e-120;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAFPKKQLGLVAATITPTMTENGESVIGQYDYLKQGVKNIFVNGTIGGLSLSV 60
Db 1 MAFPKKQLGLVAATITPTMTENGESVIGQYDYLKQGVKNIFVNGTIGGLSLSV 60
QY 61 SERQVAEWTGKOKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPEWT 120
Db 61 SERQVAEWTGKOKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPEWT 120
QY 121 KDILINFLKEVAAAAPALPFYVYHHPALTGKIRABELLDGILDKIPTFQGLKESDIDL 180
Db 121 KDILINFLKEVAAAAPALPFYVYHHPALTGKIRABELLDGILDKIPTFQGLKESDIDL 180
QY 181 DFGQCVQDNRQQQAFALFGVDEQLLSALVMGATGAVGSFVSRDLSTLLSN 230
Db 181 DFGQCVQDNRQQQAFALFGVDEQLLSALVMGATGAVGSFVSRDLSTLLSN 230

RESULT 3
AAB84681
ID AAB84681 standard; protein; 230 AA.
XX AAB84681;
AC AAB84681;
XX
DT 17-SEP-2001 (first entry)
XX
DE Amino acid sequence of a human aldolase.
XX Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;
KW cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;
KW sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;
KW vaccine.
XX
OS Homo sapiens.
XX WO200142492-A1.
PN

XX PD 14-JUN-2001.
XX PF 07-DEC-2000; 2000WO-US033136.
XX PF 09-DEC-1999; 99US-0169839P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (UJJO) UNIV JOHNS HOPKINS.
PA (UTEM) UNIV TEMPLE.
PA (UWY-) UNIV WYOMING.
XX
XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K;
PI Jarvis D;
XX
XX WPI; 2001-441575/47.
DR N-PSDB; AAH28456.
XX
XX Cells producing cytidine monophosphate-sialic acid and sialylated
PT glycoprotein above endogenous levels for production of vaccines and
PT therapeutics.
XX
XX Claim 18; Fig 28; 182pp; English.
XX
XX The specification describes a method for manipulating carbohydrate
CC processing pathways in cells of interest. The methods are used to
CC manipulate multiple pathways involved with the sialylation reaction by
CC using recombinant DNA technology and substrate feeding approaches to
CC enable the production of sialylated glycoproteins in the cells. The
CC sialylation process involves the post-translational addition of the donor
CC substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific
CC acceptor carbohydrate. The cells express at least one enzyme, selected
CC from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,
CC CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The
CC cells are useful for producing complex sialylated glycoproteins in cells
CC of interest, especially insect cells. Glycoproteins containing sialylated
CC oligosaccharides are useful as vaccines, therapeutics and diagnostic
CC tools. Cells producing complex sialylated glycoproteins are useful for
CC enhancing the value of heterologous expression systems and increasing the
CC application of heterologous cell expression products as vaccines,
CC therapeutics and diagnostic tools as well as increasing the variety of
CC heterologous proteins that can be produced and lowering biotechnology
CC production costs. The present sequence represents a human aldolase, which
CC is used in the method of the invention
XX
XX Sequence 230 AA;
SQ
Query Match 100.0%; Score 1159; DB 4; Length 230;
Best Local Similarity 100.0%; Pred. No. 8.1e-120;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAFPKKLOGLVAATITPMTENGINFVSIQGYVDYLKKEQGVKNIFVNGTTGEGLSLV 60
DB 1 MAFPKKLOGLVAATITPMTENGINFVSIQGYVDYLKKEQGVKNIFVNGTTGEGLSLV 60
QY 61 SERQVAEEWYTKGDKLDQVLIHVAGLSKESQELAHAAEIGADGIAVIAPIFLKPT 120
DB 61 SERQVAEEWYTKGDKLDQVLIHVAGLSKESQELAHAAEIGADGIAVIAPIFLKPT 120
QY 121 KDILINFLKEVAAAPALPFYTHIPALTKVIRAEELLDGILDKIPTFQGLKESDITLL 180
DB 121 KDILINFLKEVAAAPALPFYTHIPALTKVIRAEELLDGILDKIPTFQGLKESDITLL 180
QY 181 DFGQCVQDNRQQQPAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLN 230
DB 181 DFGQCVQDNRQQQPAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLN 230
RESULT 4
AAO26544
ID AAO26544
XX AAO26544 standard; protein; 230 AA.
AC AAO26544;

XX DT 06-MAR-2003 (first entry)
XX DE Human aldolase protein.
XX KW Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;
KW GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)NAC; sialic acid;
KW synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;
KW transporter; sialylated glycoprotein; human; aldolase.
XX OS Homo sapiens.
XX US2002142386-A1.
XX 03-OCT-2002.
XX 16-AUG-2001; 2001US-00930440.
XX 02-MAR-1999; 99US-0122582P.
XX 08-DEC-1999; 99US-0169624P.
XX 25-AUG-2000; 2000US-0227579P.
XX (BETE/) BETENBAUGH M J.
PA (LAWR/) LAWRENCE S.
PA (LEEY/) LEE Y C.
XX (COLE/) COLEMAN T A.
XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;
XX WPI; 2003-102519/09.
XX N-PSDB; AAL53991.
XX Manipulating glycoprotein production in insect cell, involves enhancing
PT expression of enzymes involved in carbohydrate processing pathway such as
PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.
XX
XX Claim 18; Fig 28; 88pp; English.
XX
XX The invention relates to a novel method for manipulating glycoprotein
CC production in an insect cell comprising enhancing expression of an
CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
CC catalysing conversion of UDP-GlcNAc to mannose (Man)NAC, sialic acid
CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)
CC synthetase or CMP-SA transporter, where the expression of each enzyme is
CC enhanced to above endogenous levels. The novel method is useful for
CC manipulating glycoprotein production in an insect cell. Further methods
CC of the invention are useful for producing sialylated glycoproteins. The
CC sialylated glycoprotein produced by the above mentioned methods are
CC useful as pharmaceutical compositions, vaccines, diagnostics and
CC therapeutics. This sequence represents the human aldolase protein of the
CC invention
XX
XX Sequence 230 AA;
SQ
Query Match 100.0%; Score 1159; DB 6; Length 230;
Best Local Similarity 100.0%; Pred. No. 8.1e-120;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAFPKKLOGLVAATITPMTENGINFVSIQGYVDYLKKEQGVKNIFVNGTTGEGLSLV 60
DB 1 MAFPKKLOGLVAATITPMTENGINFVSIQGYVDYLKKEQGVKNIFVNGTTGEGLSLV 60
QY 61 SERQVAEEWYTKGDKLDQVLIHVAGLSKESQELAHAAEIGADGIAVIAPIFLKPT 120
DB 61 SERQVAEEWYTKGDKLDQVLIHVAGLSKESQELAHAAEIGADGIAVIAPIFLKPT 120
QY 121 KDILINFLKEVAAAPALPFYTHIPALTKVIRAEELLDGILDKIPTFQGLKESDITLL 180
DB 121 KDILINFLKEVAAAPALPFYTHIPALTKVIRAEELLDGILDKIPTFQGLKESDITLL 180
QY 181 DFGQCVQDNRQQQPAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLN 230
DB 181 DFGQCVQDNRQQQPAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLN 230

RESULT 5
 ABB04624
 ID ABB04624 standard; protein; 284 AA.
 AC ABB04624;
 XX
 DT 04-MAR-2002 (first entry)
 XX
 DE Human dihydrodipyrindine synthetase 31 protein SEQ ID NO:2.
 XX
 KW Human; dihydrodipyrindine synthetase 31; malignant tumour; nosohaemia;
 KW HIV infection; immunological disease; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN CN1307105-A.
 XX
 PD 08-AUG-2001.
 XX
 PF 28-JAN-2000; 2000CN-00111575.
 XX
 PR 28-JAN-2000; 2000CN-00111575.
 XX
 PA (BODA-) BODAO GENE TECH CO LTD SHANGHAI.
 XX
 PI Mao Y, Xie Y;
 XX
 DR WPI; 2002-026890/34.
 DR N-PSDB; ABA24275.
 XX
 PT New polypeptide for treating malignant tumor and HIV infection, comprises
 PT the polypeptide-dihydrodi pyridine synthetase 31 and polynucleotide for
 PT coding said polypeptide.
 XX
 PS Claim 1; Page 26-27 (Disclosure); 34pp; Chinese.
 XX
 CC The present sequence represents human dihydrodipyrindine synthetase 31
 CC (1). (1) can be used in the treatment of various diseases, such as
 CC malignant tumour, nosohaemia, HIV infection, immunological diseases and
 CC inflammations. The present invention also describes the antagonist
 CC resisting (1) and its treatment effect
 CC
 SQ Sequence 284 AA;
 Query Match 95.3%; Score 1104.5; DB 5; Length 284;
 Best Local Similarity 91.0%; Pred. No. 1.2e-113;
 Matches 222; Conservative 5; Mismatches 5; Indels 15; Gaps 1;
 QY 1 MAPPKKLGQGLVAATITPTMTENGSEINFSVIGQVDYLVKEQGVKNIFVNGTTGEGLSLV 60
 DB 1 MAPPKKLGQGLVAATITPTMTENGSEINFSVIGQVDYLVKEQGVKNIFVNGTTGEGLSLV 60
 QY 61 SERQVAEEWVTGKDKLDQVITHVGALSLKESQELAQHAAEIGADGIAVIAPFELKPWT 120
 DB 61 SERQVAEEWVTGKDKLDQVITHVGALSLKESQELAQHAAEIGADGIAVIAPFELKPWT 120
 QY 121 KQILINFLKEVAAAAPALPFPYYHHPALTGVKIRAEELDGLDKIPTFGQLKFSPTDLL 180
 DB 121 KQILINFLKEVAAAAPALPFPYYHHPALTGVKIRAEELDGLDKIPTFGQLKFSPTDLL 180
 QY 181 DFGQCVQDNRQQQFAFLFGVDEQLLSALVMGATGAVGS-----FVSRDLS 225
 DB 181 DFGQCVQDNRQQQFAFLFGVDEQLLSALVMGATGAVGS-----FVSRDLS 225
 QY 226 TLIS 229
 DB 241 LAIN 244
 RESULT 6
 ABB013069
 ID ABB013069 standard; protein; 299 AA.
 AC ABB013069;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 26961.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO2001164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US004927.
 XX
 PR 28-FEB-2000; 2000US-00515126.
 PR 18-MAY-2000; 2000US-00577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI93000.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.
 XX
 PS Claim 20; SEQ ID NO 26961; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA03910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 299 AA;
 Query Match 95.3%; Score 1104.5; DB 4; Length 299;
 Best Local Similarity 91.0%; Pred. No. 1.3e-113;
 Matches 222; Conservative 5; Mismatches 5; Indels 15; Gaps 1;
 QY 1 MAPPKKLGQGLVAATITPTMTENGSEINFSVIGQVDYLVKEQGVKNIFVNGTTGEGLSLV 60
 DB 6 MAPPKKLGQGLVAATITPTMTENGSEINFSVIGQVDYLVKEQGVKNIFVNGTTGEGLSLV 65
 QY 61 SERQVAEEWVTGKDKLDQVITHVGALSLKESQELAQHAAEIGADGIAVIAPFELKPWT 120
 DB 66 SERQVAEEWVTGKDKLDQVITHVGALSLKESQELAQHAAEIGADGIAVIAPFELKPWT 125
 QY 121 KQILINFLKEVAAAAPALPFPYYHHPALTGVKIRAEELDGLDKIPTFGQLKFSPTDLL 180
 DB 126 KQILINFLKEVAAAAPALPFPYYHHPALTGVKIRAEELDGLDKIPTFGQLKFSPTDLL 185
 QY 181 DFGQCVQDNRQQQFAFLFGVDEQLLSALVMGATGAVGS-----FVSRDLS 225
 DB 186 DFGQCVQDNRQQQFAFLFGVDEQLLSALVMGATGAVGS-----FVSRDLS 245
 QY 226 TLIS 229

Db	246 LALN 249	Matches 222; Conservative 2; Mismatches 5; Indels 15; Gaps 1;
RESULT 7		
AA40417		
ID	AA40417 standard; protein; 320 AA.	
XX		
AC	AA40417;	
XX		
DT	26-NOV-1999 (first entry)	
XX		
DE	A human N-acetylneuraminase lyase (HNANL) protein.	
XX		
KW	Human; N-acetylneuraminase lyase; HNANL; N-acetylneuraminase aldolase;	
KW	acetylneuraminase-pyruvate lyase; sialic acid; degradation;	
KW	energy metabolism; sialic acid metabolism; gangliosidosis; Tay-Sachs;	
KW	beta-hexosaminidase deficiency; beta-galactosidase deficiency; cancer;	
KW	autoimmune disorder; acquired immune deficiency syndrome; AIDS;	
KW	Addison's disease; Crohn's disease; diabetes mellitus; atherosclerosis;	
KW	asthma.	
OS	Homo sapiens.	
XX		
PN	US5962302-A.	
XX		
PD	05-OCT-1999.	
XX		
PF	20-FEB-1998; 98US-00027013.	
XX		
PR	20-FEB-1998; 98US-00027013.	
XX		
PA	(INCY-) INCYTE PHARM INC.	
XX		
PI	Hillman JL, Corley NC, Shah P;	
XX		
DR	WPI; 1999-571273/48.	
DR	N-PSDB; AA222092.	
XX		
PT	Nucleic acids encoding human N-acetylneuraminase lyase useful for	
PT	treating GM1 and GM2 gangliosidosis (such as Tay-Sachs/beta-	
PT	hexosaminidase deficiency and beta-galactosidase deficiency).	
XX		
PS	Claim 1; Fig 1A-F; 31pp; English.	
XX		
CC	The present sequence represents a human N-acetylneuraminase lyase (HNANL)	
CC	protein. HNANL (also called N-acetylneuraminase aldolase and	
CC	acetylneuraminase-pyruvate lyase) is an enzyme that catalyzes the	
CC	conversion of N-acetylneuraminic acid (sialic acid) in pyruvate and N-	
CC	acetyl-D-mannosamine (and the reverse reaction). In bacteria HNANL is	
CC	involved in energy metabolism as the degradation of sialic acid produces	
CC	usable carbon and energy sources. The expression of sialic acid is	
CC	regulated by the level of sialic acid. HNANL is found in a number of	
CC	mammalian tissues and is a key enzyme in the metabolism of sialic acid.	
CC	HNANL polypeptides and polynucleotides may be used in the diagnosis,	
CC	prevention and treatment of disorders associated with defective sialic	
CC	acid metabolism, e.g. gangliosidosis (such as Tay-Sachs/beta-	
CC	hexosaminidase deficiency and beta-galactosidase deficiency). They may	
CC	also be used to study the expression and function of HNANL proteins and	
CC	their role in sialic acid metabolism. The HNANL proteins may then be used	
CC	as antigens in the production of antibodies and in assays to identify	
CC	modulators of HNANL expression and activity. These antagonists of HNANL	
CC	may be used to down regulate HNANL expression and activity which may prove	
CC	useful in the treatment of disorders associated with over expression and	
CC	activity of HNANL such as cancers (e.g. leukemia, melanoma and cancers of	
CC	the lung, brain, prostate, ovary and pancreas) and autoimmune disorders	
CC	(e.g. acquired immune deficiency syndrome (AIDS), Addison's disease,	
CC	Crohn's disease, diabetes mellitus, atherosclerosis and asthma)	
XX		
SQ	Sequence 320 AA;	
Query Match	95.3%; Score 1104.5; DB 2; Length 320;	
Best Local Similarity	91.0%; Pred. No. 1.5e-113;	
QY	1 MAPEKKKLGVAATITPMTENGESVIGQYVDYLVKQGVKNI FVNGTTGEGLSLV 60	
DB	1 MAPEKKKLGVAATITPMTENGESVIGQYVDYLVKQGVKNI FVNGTTGEGLSLV 60	
QY	61 SERQVAEEWTKGDKLDQVLIHVGALSKESELAQHAAEIGADGIAVIAPEFLKPT 120	
DB	61 SERQVAEEWTKGDKLDQVLIHVGALSKESELAQHAAEIGADGIAVIAPEFLKPT 120	
QY	121 KDILINFLKEVAAPALPFVYHIPALTGVKIPAEELLDGILDKIPTFQSKESDIDL 180	
DB	121 KDILINFLKEVAAPALPFVYHIPALTGVKIPAEELLDGILDKIPTFQSKESDIDL 180	
QY	181 DFGQVDQNFQQQFAFLFGVDEQLLSALVMGATGAVGS-----FVSRLS 225	
DB	181 DFGQVDQNFQQQFAFLFGVDEQLLSALVMGATGAVGS-----FVSRLS 225	
QY	226 TLIS 229	
DB	241 LALN 244	
RESULT 8		
AA79037		
ID	AA79037 standard; protein; 320 AA.	
XX		
AC	AA79037;	
XX		
DT	20-JUN-2000 (first entry)	
XX		
DE	Human N-acetylneuraminase lyase (HNANL) amino acid sequence.	
XX		
KW	Cancer; autoimmune disease; diagnose; treatment; leukaemia; lymphoma;	
KW	melanoma; myeloma; teratocarcinoma; sarcoma; AIDS; Addison's disease;	
KW	allergy; asthma; atherosclerosis; bronchitis; diabetes mellitus; anaemia;	
KW	Graves' disease; irritable bowel syndrome; Myasthenia gravis;	
KW	rheumatoid arthritis; osteoarthritis; Sjogren's syndrome; infection;	
KW	human N-acetylneuraminase lyase; HNANL.	
OS	Homo sapiens.	
XX		
PN	US6030824-A.	
XX		
PD	29-FEB-2000.	
XX		
PF	01-FEB-1999; 99US-00244233.	
XX		
PR	20-FEB-1998; 98US-00027013.	
XX		
PA	(INCY-) INCYTE PHARM.	
XX		
PI	Shah P, Corley NC, Hillman JL;	
XX		
DR	WPI; 2000-223164/19.	
DR	N-PSDB; AA398699.	
XX		
PT	New human N-acetylneuraminase lyase useful in the diagnosis, treatment	
PT	and prevention of cancer and autoimmune disorders, comprises of 320 amino	
PT	acid residues.	
XX		
PS	Claim 1; Fig 1; 31pp; English.	
XX		
CC	This sequence represents a human N-acetylneuraminase lyase (HNANL) amino	
CC	acid sequence. HNANL catalyzes the conversion of N-acetylneuraminic acid	
CC	(sialic acid) into pyruvate and N-acetyl-D-mannosamine, as well as the	
CC	reverse reaction. Sialic acid is an essential constituent of gangliosides	
CC	a class of glycolipids. Gangliosides are components of the cell membrane	
CC	and act as receptors and are involved in cell cell interactions, and	
CC	signal transduction. The HNANL protein sequence, and the nucleotide	
CC	sequence encoding it are useful in the diagnosis, treatment and	
CC	prevention of cancer and autoimmune disease. The cancers diagnosed,	
CC	treated and prevented include adenocarcinoma, leukaemia, lymphoma,	

CC melanoma, myeloma, teratocarcinoma, sarcoma and in particular cancers of
 CC the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix,
 CC gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver,
 CC lung, muscle, ovary, pancreas, penis, prostate, salivary glands, spleen,
 CC skin, uterus, thymus and thyroid. The autoimmune disorders diagnosed,
 CC treated and prevented include AIDS, Addison's disease, adult respiratory
 CC distress syndrome, allergies, asthma, atherosclerosis, bronchitis,
 CC diabetes mellitus, anaemia, Graves' disease, irritable bowel syndrome,
 CC myasthenia gravis, rheumatoid arthritis, osteoarthritis, Sjogren's
 CC syndrome, viral, bacterial, fungal, parasitic, protozoal and helminthic
 CC infections
 XX
 SQ Sequence 320 AA;

Query Match 95.3%; Score 1104.5; DB 3; Length 320;
 Best Local Similarity 91.0%; Pred. No. 1.5e-113;
 Matches 222; Conservative 2; Mismatches 5; Indels 15; Gaps 1;
 QY 1 MAFPKKIQGLVAATITPTMTENGESVIGQYVDYLVKEQVKNIFVNGTTGEGLSLV 60
 Db 1 MAFPKKIQGLVAATITPTMTENGESVIGQYVDYLVKEQVKNIFVNGTTGEGLSLV 60
 QY 61 SERQVABEWTKGDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPT 120
 Db 61 SERQVABEWTKGDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPT 120
 QY 121 KDILINFLKEVAAPALPFYVYHPALTGVKIRAEELLDGILDKIPTFGKFSDDTL 180
 Db 121 KDILINFLKEVAAPALPFYVYHPALTGVKIRAEELLDGILDKIPTFGKFSDDTL 180
 QY 181 DFGQVQNRQQAFAFLFGVDEQLLSALVMGATGAVGS-----FVSRDL 225
 Db 181 DFGQVQNRQQAFAFLFGVDEQLLSALVMGATGAVGS-----FVSRDL 225
 QY 226 TLIS 229
 Db 241 LALN 244

RESULT 9
 ABB97988
 ID ABB97988 standard; protein; 301 AA.
 AC ABB97988;
 XX
 DT 02-OCT-2002 (first entry)
 XX
 DE Polyadenylate binding protein 33.11.
 XX
 KW Polyadenylate binding 33.11; embryo development deformity; tumour;
 KW protein metabolism disorder.
 XX
 OS Unidentified.
 XX
 PN CN1340520-A.
 XX
 PD 20-MAR-2002.
 XX
 PF 31-AUG-2000; 2000CN-00119821.
 XX
 PF 31-AUG-2000; 2000CN-00119821.
 XX
 PR {BODE-} BODE GENE DEV CO LTD SHANGHAI.
 XX
 PA Mao Y, Xie Y;
 XX
 PI WPI; 2002-436414/47.
 DR N-PSDB; ABN86096.
 XX
 PT Polypeptide-polyadenylate binding 33.11 and polynucleotide for coding it.
 XX
 PS Claim 1; Page 29 (disclosure); 35pp; Chinese.
 XX

CC The invention relates to a novel polyadenylate binding protein 33.11, the
 CC polynucleotide encoding it, and the process for preparing the polypeptide
 CC by DNA recombination techniques. The polypeptide is used in treating
 CC diseases such as embryo development deformity, tumour and protein
 CC metabolism disorder. The current sequence represents the polyadenylate
 CC binding protein 33.11 of the invention
 XX
 SQ Sequence 301 AA;

Query Match 62.0%; Score 718.5; DB 5; Length 301;
 Best Local Similarity 85.8%; Pred. No. 9.4e-71;
 Matches 145; Conservative 4; Mismatches 5; Indels 15; Gaps 1;
 QY 76 DKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPTKDLINFLKEVAAA 135
 Db 57 ERLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPTKDLINFLKEVAAA 116
 QY 136 PALPFYVYHPALTGVKIRAEELLDGILDKIPTFGKFSDDTLDFGQVQNRQQA 195
 Db 117 PALPFYVYHPALTGVKIRAEELLDGILDKIPTFGKFSDDTLDFGQVQNRQQA 176
 QY 196 FLFGVDEQLLSALVMGATGAVGS-----FVSRDLSTLIS 229
 Db 177 FLFGVDEQLLSALVMGATGAVGS-----FVSRDLSTLIS 225

RESULT 10
 AAG81340
 ID AAG81340 standard; protein; 147 AA.
 AC AAG81340;
 XX
 DT 10-SEP-2001 (first entry)
 XX
 DE Human AFP protein sequence SEQ ID NO:198.
 XX
 KW Human; secreted protein; secretion; bacterial cell; fungal cell;
 KW eukaryotic cell; fusion protein; maltose binding protein;
 KW immunoglobulin constant region; polyhistidine tag.
 XX
 OS Homo sapiens.
 XX
 PN WO200129221-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000WO-US029052.
 XX
 PR 20-OCT-1999; 99US-0150712P.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Conklin DC, Yee DP;
 XX
 PN WPI; 2001-300340/31.
 DR N-PSDB; AAH52191.
 XX
 PT Isolated polypeptide for directing secretion of proteins of interest from
 PT a host cell including, e.g. bacteria, includes contiguous amino acid
 PT residues of polypeptide with specified amino acids.
 XX
 PS Claim 1; Page 348-349; 617pp; English.
 XX
 CC AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242
 CC to AAG81453. The secreted proteins can be used for directing the
 CC secretion of proteins of interest from a host cell including bacteria,
 CC fungal cells, and cultured higher eukaryotic cells. The present invention
 CC also describes fusion proteins, where a secreted protein of the invention
 CC is operably linked via a peptide bond or peptide linker to a second
 CC protein selected from the group consisting of maltose binding protein, an
 CC immunoglobulin constant region, a polyhistidine tag and a peptide given
 CC in AAG81453
 XX

```
SQ Sequence 147 AA;
Query Match 50.2%; Score 675; DB 4; Length 147;
Best Local Similarity 93.2%; Pred. No. 2.2e-66;
Matches 138; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 1 MAFPKKLGVAATITPMTENGESVIGQYVDYLVKEQGVKNIFVNGTGGGLSLSV 60
DQ 1 MAFPKKLGVAATITPMTENGESVIGQYVDYLVKEQGVKNIFVNGTGGGLSLSV 60
QY 61 SERQVAEEWTKGKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPIFLKQWT 120
DQ 61 SERQVAEEWTKGKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPIFLKQWT 120
QY 121 KDILINFLKEVAAPALPFFYYHIPAL 148
DQ 121 KDILINFLKEVAAPALPFFYYHIPAL 148

RESULT 11
AAM90590
ID AAM90590 standard; protein; 141 AA.
XX AC AAM90590;
XX DT 07-NOV-2001 (first entry)
XX DE Human immune/haematopoietic antigen SEQ ID NO:18183.
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KY cytosolic; gene therapy; vaccine; metastasis.
XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001354.
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186309P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0195123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227609P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
```

```

PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259676P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX N-PSDB; AAK63371.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Claim 11; SEQ ID NO 18183; 3071bp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK62170 to AAK61921. (I) have cytotoxic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention
XX
XX Sequence 141 AA;
XX
XX Query Match 51.7%; Score 599; DB 4; Length 141;
XX Best Local Similarity 100.0%; Pred. No. 5.6e-58;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 102 EIGADGIAVAPFLKFWTKDIINFLKEVAAPALPFYVYHHPALTGKIRAEELLOG 161
XX
XX 3 EIGADGIAVAPFLKFWTKDIINFLKEVAAPALPFYVYHHPALTGKIRAEELLOG 62
XX
XX 162 ILDKIPTFQGLKFSQDQNRQQQVDFGVDEQLLSALVMGATGAVGS 218
XX

```

```

Db 63 ILDKIPTFQGLKFSQDQNRQQQVDFGVDEQLLSALVMGATGAVGS 119
RESULT 12
AAR74749
ID AAR74749 standard; protein; 297 AA.
XX
XX AAR74749;
XX
XX 10-JAN-1996 (first entry)
XX
XX Escherichia coli N-acetylneuraminic acid aldolase.
XX
XX Escherichia coli; N-acetylneuraminic acid aldolase; sialic acid;
XX recombinant production; carcinoma diagnosis.
XX
XX Escherichia coli.
XX JP07107968-A.
XX
XX 25-APR-1995.
XX
XX 14-OCT-1993; 93JP-00256903.
XX
XX 14-OCT-1993; 93JP-00256903.
XX
XX (TOYM) TOYOBOKK.
XX
XX WPI; 1995-130169/25.
XX N-PSDB; AAK92410.
XX
XX Recombinant proctn. of N-acetylneuraminic acid aldolase - in gram-negative
XX bacterium other than Escherichia sp., useful for measuring change in
XX sialic acid content for diagnosis of carcinoma.
XX
XX Claim 1; Page 7-8; 13pp; Japanese.
XX
XX AAK92410 encodes AAR74749 Escherichia coli N-acetylneuraminic acid
XX aldolase, which can be recombinantly produced in other Gram negative
XX bacteria. The enzyme is useful for measuring changes in sialic acid
XX content, and is therefore useful in the diagnosis of carcinomas
XX
XX Sequence 297 AA;
XX
XX Query Match 23.7%; Score 274.5; DB 2; Length 297;
XX Best Local Similarity 34.3%; Pred. No. 1.8e-21;
XX Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;
XX
XX 8 LQGLVAATITPMTENGSEINFSVIGQVYDLVKQGVKNIFVNGTGTGGLSLVSERRQ-- 65
XX
XX 5 LRGVMAALITPFPQQQALDKASRLRLVQFNI--QQGIDGLYVGSGTGEAFVCSLSEREQVL 63
XX
XX 66 --VAREWTKGDKLDQVIVHVGALSLSQESQELAQHAAEIGADGIAVAPFLKFWTKDI 123
XX
XX 64 EIVARE--AKGKIKL--IAHVGCVSTAESQQAASAKRYGFDVAVSATVPFY--PFSFEE 117
XX
XX 124 LINFLEKVAAPALPFYVYHHPALTGKIRAEELLOGIILDKIPTFQGLKFSQDQNRQQQ 183
XX
XX 118 HCDHYRAIDISAGLPMVYVNIIPALSGVKLTLDQI--NTLVTPGVGALKQTSGLDQME 175
XX
XX 184 QCVDNRQQQVDFGVDEQLLSALVMGATGAVGS 218
XX
XX 176 QI---RREHPDIVLYNGYDEIFASGLLAGADGGIGS 208
XX
RESULT 13
AD90650
ID AD90650 standard; protein; 297 AA.
XX
XX AD90650;
XX
XX 29-JAN-2004 (first entry)
XX

```

DE Escherichia coli protein. SEQ ID NO:3.
 XX microorganism; N-acetylneuraminic acid; Escherichia coli.
 KW Escherichia coli.
 XX WC2003072783-A1.
 XX 04-SEP-2003.
 XX 28-FEB-2003; 2003WO-JP0023332.
 XX 28-FEB-2002; 2002JP-00053696.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX Ishikawa M, Koizumi S;
 XX WPI; 2003-721776/68.
 XX N-PSDB; ADD90651.
 XX Microbial production of N-acetylneuraminic acid for use e.g. in chemical
 PT synthesis or as pharmaceutical intermediates.
 XX Disclosure; SEQ ID NO 3; 52pp; Japanese.
 XX The present invention describes a process for producing N-
 CC acetylneuraminic acid by culturing a microorganism capable of producing N-
 CC acetylneuraminic acid but with no or lower activity in decomposing N-
 CC acetylneuraminic acid as compared with the wild-type strain, in a culture
 CC medium to accumulate the product for isolation. The method can be used
 CC for the production of N-acetylneuraminic acid for use e.g. in chemical
 CC synthesis or as pharmaceutical intermediates. The process is efficient
 CC and economical, in which a microorganism used has no or lower activity in
 CC decomposing N-acetylneuraminic acid as compared with the wild-type
 CC strain, from cheap starting materials and with simple operation. The
 CC present sequence represents an Escherichia coli protein, which is used in
 CC the exemplification of the present invention.
 XX
 XX Sequence 297 AA;

Query Match 23.7%; Score 274.5; DB 7; Length 297;
 Best Local Similarity 34.3%; Pred. No. 1.8e-21;
 Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;
 QY 8 LQGLVAATITMTENGSENFVIGQYVDYLVKEQGVKAFVNGTGTGEGLSLSVSRQ-- 65
 DB 5 LRGVWAALITPDQQALDKASRLRVFNFI--QGGIDGLYVGGSTGEAFVQSLAERQ 63
 QY 66 --VAPSWTKGDKLDQVVIHVGLSLSKESQELAQHAAEIGADGIAVIAPFLKPTWKDI 123
 DB 64 EIVAE--AKGKIL--LAHVGVSTAESQQLAASAKRYGFDVA SVATVPFY--PFSFE 117
 QY 124 LINFKEVAAPALPFFYYHIPALTGVKIRAEELLDGILDKIPTFGKLSVSRQ 183
 DB 118 HCDYRAIDSDAGLEWVYVNIIPALSGVKLTLDQI--NTLVTLPGVGALKQTSGLDYQME 175
 QY 184 QCVQNRQQQFAFLF-GVDEQLLSALVWGATGAVGS 218
 DB 176 QI---RRHPDLVLYNGYDEIFASGLLAGADGGIGS 208

RESULT 14
 AAB85682
 ID AAB85682 standard; protein; 297 AA.
 XX
 AC AAB85682;
 XX
 XX 29-OCT-2001 (first entry)
 XX Salmonella virulent protein.
 XX Salmonella; virulent; medicament; infection; Gram-negative bacteria;

KW gastro-enteritis; veterinary; antimicrobial; antibacterial; vaccine;
 XX antiinflammatory.
 OS Salmonella typhimurium.
 PN WC200157075-A2.
 XX 09-AUG-2001.
 XX 02-FEB-2001; 2001WO-GB000449.
 XX 03-FEB-2000; 2000GB-00003552.
 XX 03-FEB-2000; 2000GB-00003554.
 XX 03-FEB-2000; 2000GB-00003555.
 XX 03-FEB-2000; 2000GB-00003556.
 XX 03-FEB-2000; 2000GB-00003557.
 XX 03-FEB-2000; 2000GB-00003558.
 XX 03-FEB-2000; 2000GB-00003559.
 XX 03-FEB-2000; 2000GB-00003560.
 XX 03-FEB-2000; 2000GB-00003561.
 XX 03-FEB-2000; 2000GB-00003562.
 XX 03-FEB-2000; 2000GB-00003563.
 XX 03-FEB-2000; 2000GB-00003564.
 XX 20-SEP-2000; 2000GB-00023059.
 XX 20-SEP-2000; 2000GB-00023060.
 XX (MICR-) MICROSCIENCE LTD.
 PA Chatfield SN;
 XX WPI; 2001-476280/51.
 XX N-PSDB; AAB47030.
 XX Novel peptide encoded by a virulence gene of Salmonella typhimurium is
 PT useful for the treatment or prevention of a condition associated with
 PT infection by Salmonella or Gram-negative bacteria, e.g., gastro-
 PT enteritis.
 XX
 XX Claim 4; Page 39-41; 63pp; English.

The invention provides Salmonella typhimurium virulent proteins (AAB85675
 -688) and nucleotide sequences (AAH47025-036) encoding the virulent
 CC proteins. The proteins can be expressed by standard recombinant
 CC methodology. The virulent genes and proteins and vaccines and antibodies
 CC raised against the proteins may be used for the manufacture of a
 CC medicament for use in the treatment or prevention of a condition
 CC associated with infection by Salmonella or Gram-negative bacteria, e.g.,
 CC gastro-enteritis. They may also be used for veterinary treatment as well
 CC as in a screening assay for the identification of an antimicrobial drug.
 CC The present sequence represents a Salmonella virulent protein

Sequence 297 AA;

Query Match 23.5%; Score 272.5; DB 4; Length 297;
 Best Local Similarity 31.9%; Pred. No. 2.9e-21;
 Matches 68; Conservative 43; Mismatches 95; Indels 7; Gaps 5;
 QY 6 KKLQGLVAATITMTENGSENFVIGQYVDYLVKEQGVKAFVNGTGTGEGLSLSVSRQ 65
 DB 3 KALQGVWAALITPDHQQQLDSESLRLRVFNFI--QGGIDGLYVGGSTGEAFVQSLAERQ 61
 QY 66 VAEWVTKGDKLDQVVIHVGLSLSKESQELAQHAAEIGADGIAVIAPFLKPTWKDI 125
 DB 62 VLEIVAEAKGI-TLIAHVGVSTAESQQLAASAKRYGFDVA SVATVPFY--PFSFEHC 119
 QY 126 NFLKEVAAPALPFFYYHIPALTGVKIRAEELLDGILDKIPTFGKLSVSRQ 185
 DB 120 DHYRAIDSDAGLEWVYVNIIPALSGVKLTLDQI--NTLVTLPGVGALKQTSGLDYQME 177
 QY 186 VTCNRQQQFAFLF-GVDEQLLSALVWGATGAVGS 218
 DB 178 --RRHPDLVLYNGYDEIFASGLLAGADGGIGS 208

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:00:50 ; Search time 8.7363 Seconds
(without alignments)
2532.427 Million cell updates/sec

Title: US-09-930-440B-2

Perfect score: 1159

Sequence: 1 MAPPKKLGGLVAATTPMT.....GATGAVGSFVSRDLSTLSN 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	443	38.2	297	2	AG0367
2	430	37.1	298	2	probable N-acetylneuraminat
3	274.5	23.7	297	1	WZECN
4	274.5	23.7	297	2	N-acetylneuraminat
5	274.5	23.7	297	2	N-acetylneuraminat
6	272.5	23.5	297	2	AB0908
7	256	22.1	305	2	B95195
8	252.5	21.8	305	2	B95154
9	251.5	21.7	294	2	H75033
10	242.5	20.9	293	2	E89796
11	241	20.8	313	2	A84203
12	240.5	20.8	293	2	G64050
13	237.5	20.5	299	2	A92878
14	232	20.0	302	2	D90489
15	231.5	20.0	283	2	A6207
16	231.5	20.0	289	2	E90397
17	230.5	19.9	287	2	C71135
18	224.5	19.4	291	2	H81878
19	222.5	19.2	301	1	C4626
20	222	19.2	294	2	E70398
21	218.5	18.9	319	1	S56523
22	218	18.8	309	2	D64752
23	217.5	18.8	291	2	D81141
24	208	17.9	294	2	AH2265
25	208	17.9	302	2	B90485
26	205.5	17.7	294	2	AG2702
27	205.5	17.7	294	2	G97484
28	204.5	17.6	292	2	G83949
29	203.5	17.6	290	2	E46665

30	201.5	17.4	289	2	E64330	dihydrodipicolinat
31	201	17.3	301	2	S76023	dihydrodipicolinat
32	200	17.3	298	2	A64058	dihydrodipicolinat
33	197.5	17.0	317	2	T36926	dihydrodipicolinat
34	197.5	17.0	333	1	E53308	mosa protein - Rhi
35	196.5	17.0	292	2	C83520	dihydrodipicolinat
36	196.5	17.0	295	2	F83867	dihydrodipicolinat
37	195	16.8	293	2	F87193	dihydrodipicolinat
38	195	16.8	299	2	T35844	dihydrodipicolinat
39	192	16.6	294	2	F87397	dihydrodipicolinat
40	189.5	16.4	307	2	AG3414	dihydrodipicolinat
41	188	16.2	300	2	H70879	dihydrodipicolinat
42	188	16.2	308	2	G90504	2-keto-3-deoxy glu
43	187.5	16.2	294	2	F72246	dihydrodipicolinat
44	186	16.0	289	2	F69363	dihydrodipicolinat
45	185.5	16.0	295	2	C89916	dihydrodipicolinat

ALIGNMENTS

RESULT 1

AG0367

Probable N-acetylneuraminat lyase (EC 4.1.3.3) [imported] - Yersinia pestis (strain CO C/Species: Yersinia pestis
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C/Accession: AG0367

R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B. deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AG0367

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-297 <KUR>

A/Cross-references: GB:AL590842; PIDN:CAC92266.1; PID:gl5980977; GSPDB:GN00175

C/Genetics:

A/Gene: YPO3024

C/Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 38.2%; Score 443; DB 2; Length 297;

Best Local Similarity 43.7%; Pred. NO. 4.9e-29;

Matches 93; Conservative 41; Mismatches 75; Indels 4; Gaps 4;

QY 6 KKLQGLVAATITPTENGESVIGQYDYLKKEGVKNIFVNGITGESLSVSRQ 65

Db 2 KKLTHIAAPHTPFDEQGEVNVPEIDIAEHLND-GVKGVYCGTGEIHCSVDKRX 60

QY 66 VAEWVTKCKDKIDQVITHVGALSLKESQELAQAAEIGADGIAVIAPFLKPTWKDILI 125

Db 61 IAEKRWVNAAGKL-SITLHTGALSIRKDAVLSHAETLDFATSAIGPCFFKGNLDELI 119

QY 126 NLFKEVAAAPALPFYVYHIPALTYGKIRAEELLDGILDKIPTFQGLKSDTDLDPGQC 185

Db 120 AYCOALAAAPSGFYVYH-SGMSGVNLDMQFLKAESKIPNLSGIKFNADLYEFQRC 178

QY 186 VQNRQOQAFIFGVDEQLSLVNWGATGAVGS 218

Db 179 L-RVSGKGFIDIPFVDEHLPGGLAVGAIGAVGS 210

RESULT 2

EB2157

probable N-acetylneuraminat lyase VC1776 [imported] - Vibrio cholerae (strain N16961 s. C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: EB2157

R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J. chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragci, I.; Sellers, I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.N. Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A:Reference number: AB2C35; MUID:2046833; PMID:10952301
 A:Accession: E82157
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-298 <HEL>
 A:Cross-references: GB:AE004255; GB:AE003852; NID:g9656299; PIDN:AAF94925.1; GSPDB:GN001
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC1776
 A:Map position: 1

Query Match 37.1%; Score 430; DB 2; Length 298;
 Best Local Similarity 44.1%; Pred. No. 5.7e-28;
 Matches 94; Conservative 39; Mismatches 76; Indels 4; Gaps 4;

Qy 6 KKLQGLVAATITPTWENGELNFSVIGQYVDYLVKEQGVKNI FVNGTTGEGLSLSVSERRQ 65
 Db 2 KKLGLIAAPHTPTKDNKNVFAADQLAELL EGVKGAVVCGTTGEGIHCSVEERKA 60

Qy 66 VAEWVTGKDKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFFLKPTWKDI 125
 Db 61 IAEWRVKAVDKLD-VILHTGALSIVDTNLTAEATLDFIATSAIGPCFFKPGSGWDLV 119

Qy 126 NLFKEVAAPALPFYVHIPALTGKIRABEILDGILDKIPTFOGLKESDITLLDFQC 185
 Db 120 EYCAQVAAASKGFYVH-SGMSGVNLDLQFLKGSQRIPNLYGAKFNADLYEYQRC 178

Qy 186 VDQNRQQQAFALF-GVDEQLLSALVMGATGAVGS 218
 Db 179 V-RVSNRRFDIPFGVDEFLPAGLAUGAVGAVGS 210

RESULT 3
 WZECN
 N-acetylneuraminatase lyase (EC 4.1.3.3) - *Escherichia coli* (strain K-12)
 A:Alternate names: N-acetylneuraminatase algolase
 C:Species: *Escherichia coli*
 C:Date: 31-Mar-1988 #sequence_revision 30-Sep-1990 #text_change 01-Mar-2002
 C:Accession: JF0002; A24109; C65114
 R:Kawakami, B.; Kudo, T.; Narahashi, Y.; Horikoshi, K.
 Agric. Biol. Chem. 50, 2155-2159, 1986
 A:Title: Nucleotide sequence of the N-acetylneuraminatase lyase gene of *Escherichia coli*.
 A:Reference number: A90021
 A:Accession: JF0002
 A:Molecule type: DNA
 A:Residues: 1-297 <KAW>
 A:Experimental source: K12, strain H3101
 R:Ohta, Y.; Watanabe, K.; Kimura, A.
 Nucleic Acids Res. 13, 8843-8852, 1985
 A:Title: Complete nucleotide sequence of the *E. coli* N-acetylneuraminatase lyase.
 A:Reference number: A93603; MUID:86093682; PMID:3909108
 A:Accession: A24109
 A:Molecule type: DNA
 A:Residues: 1-69, 'G', '71-83', 'T', '85-297' <OHT>
 A:Experimental source: K12, strain C500
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 .A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: C65114
 A:Status: preliminary; nucleic acid sequence not shown; translation: not shown
 A:Molecule type: DNA
 A:Residues: 1-297 <BLAT>
 A:Cross-references: GB:AE000402; GB:J00096; NID:g1789619; PIDN:AACT6257.1; PID:g1789620;
 A:Experimental source: strain K-12, substrain MG1655
 C:Comment: This enzyme catalyzes the conversion of N-acetylneuraminic acid to pyruvate
 C:Genetics:
 A:Gene: nanA; npl
 A:Map position: 69 min
 C:Superfamily: dihydroadipicolinate synthase
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase

F165/Active site: Lys (covalent pyruvate-binding); #status predicted
 Query Match 23.7%; Score 274.5; DB 1; Length 297;
 Best Local Similarity 34.3%; Pred. No. 3.7e-15;
 Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;

Qy 8 LQGLVAATITPTWENGELNFSVIGQYVDYLVKEQGVKNI FVNGTTGEGLSLSVSERRQ-- 65
 Db 5 LRGVAAALTPDQQQALDKASRLRVQFNI--QQGIDGLYGGSTGEAFVQSLSREQVL 63

Qy 66 --VAEWTGKDKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFFLKPTWKDI 123
 Db 64 EIVAE--AKGKIL--IAHVGCVSTAESQQLAASAKRYGFDVAVTPFY--PFSPEE 117

Qy 124 LINFKEVAAPALPFYVHIPALTGKIRABEILDGILDKIPTFOGLKESDITLLDFG 183
 Db 118 HCDHYRAIDSADGLPMVWVNI PALSGVKLTLDQI--NTLVLPFGVGALKQTSGLYQME 175

Qy 184 QCVDNRRQQQAFALF-GVDEQLLSALVMGATGAVGS 218
 Db 176 QI---RREHPDLVLYNGYDEIFASGLLAGAGGGIGS 208

RESULT 4
 E85986
 N-acetylneuraminatase lyase (EC 4.1.3.3) - *Escherichia coli* (strain O157:H7, substrain EI
 C:Species: *Escherichia coli*
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
 C:Accession: E85986
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
 Nature 409, 523-533, 2001
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: E85986
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-297 <STO>
 A:Cross-references: GB:AE005174; NID:gl2517843; PIDN:AA058353.1; GSPDB:GN00145; UWGP:Z4
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: nanA
 C:Superfamily: dihydroadipicolinate synthase
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 23.7%; Score 274.5; DB 2; Length 297;
 Best Local Similarity 34.3%; Pred. No. 3.7e-15;
 Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;

Qy 8 LQGLVAATITPTWENGELNFSVIGQYVDYLVKEQGVKNI FVNGTTGEGLSLSVSERRQ-- 65
 Db 5 LRGVAAALTPDQQQALDKASRLRVQFNI--QQGIDGLYGGSTGEAFVQSLSREQVL 63

Qy 66 --VAEWTGKDKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFFLKPTWKDI 123
 Db 64 EIVAE--AKGKIL--IAHVGCVSTAESQQLAASAKRYGFDVAVTPFY--PFSPEE 117

Qy 124 LINFKEVAAPALPFYVHIPALTGKIRABEILDGILDKIPTFOGLKESDITLLDFG 183
 Db 118 HCDHYRAIDSADGLPMVWVNI PALSGVKLTLDQI--NTLVLPFGVGALKQTSGLYQME 175

Qy 184 QCVDNRRQQQAFALF-GVDEQLLSALVMGATGAVGS 218
 Db 176 QI---RREHPDLVLYNGYDEIFASGLLAGAGGGIGS 208

RESULT 5
 B91141
 N-acetylneuraminatase lyase [imported] - *Escherichia coli* (strain O157:H7, substrain RIM
 C:Species: *Escherichia coli*
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Aug-2002
 C:Accession: B91141
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic
A:Reference number: A99629; PMID:21156231; PMID:11253796

A:Accession: B91141
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA037521.1; PID:g13363571; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
C:Superfamily: dihydrodipicolinate synthase

Query Match 23.7%; Score 274.5; DB 2; Length 297;
Best Local Similarity 34.3%; Pred. No. 3.7e-15;
Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;

Qy 6 LQGLVAATITPMTENGSEINFSVIGQYVDYLKVEQGVKNIFVNGTTGEGLSLSVSERRQ-- 65

Db 5 LRGVAAALTPDQQAALDXSLRLVQFNI--QGGIDGLYVGSGTGEAFVQSLSREQVL 63

Qy 66 --VAEEWTKGKDKLDQVIHVGLSLKESQELAHAAEIGADGIAVIAPEFLKPTWKDIL 123

Db 64 EIVAE--AKGKIL--IAHVGCYSTAESQQLAASAKRYGFDVAVTPFFY--PFSFEH 117

Qy 124 LINFLEVAAPALPFYHYHIPALTGVKIRAEELDGLDKIPFGQLKFSDTLLDFG 183

Db 118 HCDHYRAIDSGDGLFMVYNIPALSGVKLTLDQI--NTLVLPGVGALKQTSGLDYQME 175

Qy 184 QCVQDNROQQFAFLF--GVDEQLLSALVMGATGAVS 218

Db 176 QI---RRHPDLVLYNGDEIFASGLLAGAGGGIGS 208

RESULT 6
AB0908
K-acetylneuraminate lyase [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhi

C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A:Note: this species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AB0908
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Comerston, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov

A:Reference number: AB0502; PMID:21534947; PMID:11677608
A:Accession: AB0908
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-297 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07856.1; PID:g16504403; GSPDB:GN00176

C:Genetics:
A:Gene: STX3520
C:Superfamily: dihydrodipicolinate synthase

Query Match 23.5%; Score 272.5; DB 2; Length 297;
Best Local Similarity 31.9%; Pred. No. 5.3e-15;
Matches 68; Conservative 43; Mismatches 95; Indels 7; Gaps 5;

Qy 6 KKQLGLVAATITPMTENGSEINFSVIGQYVDYLKVEQGVKNIFVNGTTGEGLSLSVSERRQ 65

Db 3 KALQGVAAALTPDQQAALDXSLRLVQFNI--QGGIDGLYVGSGTGEAFVQSLSREQ 61

Qy 66 VAEWTKGKDKLDQVIHVGLSLKESQELAHAAEIGADGIAVIAPEFLKPTWKDILI 125

Db 62 VLEIVAEAKGI--TLIAHVGVSTAESQQLAASAKRYGFDVAVTPFFY--PFSFEHC 119

Qy 126 NPLKEVAAPALPFYHYHIPALTGVKIRAEELDGLDKIPFGQLKFSDTLLDFGQC 185

Db 120 DHYRAIDSGDGLFMVYNIPALSGVKLTLDQI--NTLVLPGVNALKQTSGLDYQMEQI 177

Qy 186 VQDNROQQFAFLF--GVDEQLLSALVMGATGAVS 218

Db 178 --RRHPDLVLYNGDEIFASGLLAGAGGGIGS 208

RESULT 7
B95195
K-acetylneuraminate lyase, probable [imported] - *Streptococcus pneumoniae* (strain TIGR4

C:Species: *Streptococcus pneumoniae*
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: B95195
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei,
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.

A:Reference number: A95000; PMID:21357209; PMID:11463916
A:Accession: B95195
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-305 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75755.1; PID:g14973168; GSPDB:GN00164; TIGR:SP

A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI676

Query Match 22.1%; Score 256; DB 2; Length 305;
Best Local Similarity 31.2%; Pred. No. 1.3e-13;
Matches 67; Conservative 42; Mismatches 94; Indels 12; Gaps 6;

Qy 6 KKQLGLVAATITPMTENGSEINFSVIGQYVDYLKVEQGVKNIFVNGTTGEGLSLSVSERRQ 65

Db 5 KKEGVIPAFYACYDEQGEVSPERTRALVQYFI--DKGVQGLYNGSGECITQGVSDRKL 63

Qy 66 VAEWTKGKDKLDQVIHVGLSLKESQELAHAAEIGADGIAVIAPEFLKPTWKDIL 125

Db 64 ILEEVNAVAKGL--TLIAHVACNNTKSNELAHASGLVDALATPPYIFR--LPYSVA 121

Qy 126 NPLKEVAAPALPFYHYHIPALTGVKIRAE--ELLGDKIPFGQLKFSDTLLDF 182

Db 122 KYWNDISSAAPTVDYINIPQLAGVALTPSLYTEML-----KNPEVIGKSSMPVQDI 176

Qy 183 GCVQDNROQQFAFLF--GVDEQLLSALVMGATGAVS 217

Db 177 QTFVSLGGEDHIVF--NGPDEQFLGGLMGARAGIG 210

RESULT 8
B95154
K-acetylneuraminate lyase [imported] - *Streptococcus pneumoniae* (strain TIGR4)

C:Species: *Streptococcus pneumoniae*
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: B95154
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei,
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.

A:Reference number: A95000; PMID:21357209; PMID:11463916
A:Accession: B95154
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-305 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75427.1; PID:g14972811; GSPDB:GN00164; TIGR:SP

A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI329

Query Match 21.8%; Score 252.5; DB 2; Length 305;
Best Local Similarity 31.1%; Pred. No. 2.4e-13;

RESULT 12
 G64050
 N-acetylneuraminase lyase (EC 4.1.3.3) [similarity] - Haemophilus influenzae (strain Rd
 C:Species: Haemophilus influenzae
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 21-Jul-2000
 C:Accession: G64050
 R.Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
 Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.
 D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A.:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630; PMID:7542800
 A:Accession: G64050
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-293 <TIGR>
 A:CROSS-references: GB:U32700; GB:L42023; NID:g3212181; PIDN:AAC2181.4.1; PID:g1573098; T
 C:Superfamily: dihydrodipicolinate synthase
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase
 F:164/Active site: Lys (covalent pyruvate-binding) #status predicted

Query Match 20.8%; Score 240.5; DB 2; Length 293;
 Best Local Similarity 28.9%; Pred. No. 2.3e-12;
 Matches 66; Conservative 44; Mismatches 81; Indels 37; Gaps 6;
 QY 6 KKLGLVAATITPMTENGNEINFSVIGQVYDLVKEQGVKNIFVNGTTGEGLSLSVSERRQ 65
 DB 2 RDLGIFSAALLVSFNEGDTINEKGLAQIRHNIDKMKVGLYVGGSTGENFPM-STBEKE 61
 QY 66 VAEWVTGKDKLDQVLIHVGLSLSQELAQHAAEIGADGIAVIAPPLKXETKDILI 125
 DB 62 IFRIAKDEAKQI-ALLAQVGSVNLKAEVLSKATLELGVDCLSATVPFYK----- 112
 QY 126 NFEKEVAAPALPFY-----YHIALTGKVKIRAEELLDGILDKIPTFQGL 172
 DB 113 -----FSFPEIKHYVDITIAETGNMTIVSIPFLTGVMNGIEQF--GELYKNPKVLGV 163
 QY 173 KPSDITD--ILLDQGVQVQNRQQQFAELFGVDEQLLSALVMGATGAVGS 218
 DB 164 KFTAGDYLLERLKKAYPNH---LIWAGFDEWMLPAASLGVGDGATGS 207

RESULT 13
 A95878
 probable N-acetylneuraminase lyase subunit protein [imported] - Sinorhizobium meliloti
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 27-Nov-2001
 C:Accession: A95878
 R.Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: A95878
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-299 <KUR>
 A:CROSS-references: GB:AL591985; PIDN:CAC48689.1; PID:g15140161; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymB
 R.Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Leilaure,
 Hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: nanA; SWb20299
 A:Genome: plasmid
 C:Superfamily: dihydrodipicolinate synthase

Query Match 20.5%; Score 237.5; DB 2; Length 299;
 Best Local Similarity 29.8%; Pred. No. 4.1e-12;
 Matches 62; Conservative 49; Mismatches 92; Indels 9; Gaps 7;
 QY 7 KLOGLVAATITPMTENGNEINFSVIGQVYDLVKEQGVKNIFVNGTTGEGLSLSVSERRQV 66
 DB 2 KLEGIYSALLTPESESDIDRAIGALVDQVR-LGIDGVYVGGSSGEAMLOSILDERADY 60
 QY 67 AEWVTGKDKLDQVLIHVGLSLSQELAQHAAEIGADGIAVIAPPLKXETKDILI 126
 DB 61 LSDVAAAASGRLL-TLIAHVGTIATRDALRLSQAASQYQAISAPFPYDFDSRPEVMAH 119
 QY 127 FLKEVAAAAPALPFYHYHIALTGKVKIRAEELLDGILDKIPTFQGLKFSITDILLDFQCV 186
 DB 120 Y-RELADWS-ALFLIVNFPARTS-QFTLPFLVE--LLSHENIIGIKHTSDMFQLERI- 173
 QY 187 DQNRQQQFAELFGVDEQLLSALVMGATGAVGS 218
 DB 174 -RHAVPDAIVNGYDEMCLAGFANGAGGAGT 204

RESULT 14
 D90489
 dihydrodipicolinate synthase (dapA-3) [imported] - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
 C:Accession: D90489
 R.; She, Q.; Singh, R.K.; Confalonieri, P.; Zivanovic, Y.; Allard, G.; Awavez, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, I.
 arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001
 A:Description: Sulfolobus solfataricus complete genome.
 A:Reference number: A99139
 A:Accession: D90489
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-302 <KUR>
 A:CROSS-references: GB:AE006641; NID:g13816479; PIDN:NAK43171.1; GSPDB:GN00155
 C:Genetics:
 A:Gene: dapA-3

Query Match 20.0%; Score 232; DB 2; Length 302;
 Best Local Similarity 31.8%; Pred. No. 1.2e-11;
 Matches 68; Conservative 40; Mismatches 96; Indels 10; Gaps 6;
 QY 7 KLOGLVAATITPMTENGNEINFSVIGQVYDLVKEQGVKNIFVNGTTGEGLSLSVSERRQV 66
 DB 2 RLEGVVVPLITPLEDYSIDKEGLKWLVSYL-SENGVNGIFVNSVGEFASLTIDEMKLV 60
 QY 67 AEWVTGKDKLDQVLIHVGLSLSQELAQHAAEIGADGIAVIAPPLKXETKDILI 124
 DB 61 TK--IALDSKRSYSTLIIFAGASTNFTETIKLAKDFKMGVDAIVVMAPYFFKVRKELL 118
 QY 125 INFLKEVAAAAPALPFYHYHIALTGKVKIRAEELLDGILDKIPTFQGLKFSITDILLDFGQ 184
 DB 119 DHP--SMIAEKVDLPILIIYNIPLFTGIDIPV-SVKMLVSYQHSNIIGTKVTLDSLFFKK 175

RESULT 15
 A69207
 dihydrodipicolinate synthase (EC 4.2.1.52) MTH801 [similarity] - Methanobacterium therm
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 31-Mar-2000
 C:Accession: A69207
 R.; Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
 Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.
 ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct

Search completed: May 6, 2004, 09:09:41
Job time : 10.9863 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:56:04 ; Search time 6.04821 Seconds
(without alignments)
1980.112 Million cell updates/sec

Title: US-09-930-440B-2

Perfect score: 1.59

Sequence: 1 YAFPRKKLQGLVAATITPT.....GATGAGSFVSRDLSTLLSN 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	274.5	23.7	296	1 NANA_ECOLI	P06995 escherichia
2	273.5	23.6	296	1 NANA_ECOLI	Q8f358 escherichia
3	272.5	23.5	297	1 NANA_SALTI	Q8z3f0 salmonella
4	272.5	23.5	297	1 NANA_SALTI	Q8z3f0 salmonella
5	247.5	21.4	293	1 NANA_PASMU	Q9ckb0 pasteurella
6	246.5	21.3	305	1 NANA_ECOLI	Q8fdu7 escherichia
7	242.5	20.9	293	1 NANA_STAAW	Q99wt1 staphylococ
8	242.5	20.9	293	1 NANA_STAAW	Q8nvc7 staphylococ
9	241.5	20.8	292	1 NANA_LACPL	P59407 lactobacill
10	241.5	20.8	304	1 NANA_HALNI	Q9hs19 halobacteri
11	240.5	20.8	293	1 NANA_HAEIN	P44539 haemophilus
12	237.5	20.5	299	1 NANA_RHIME	Q92wp0 rhizobium m
13	234.5	20.2	300	1 DAPA_METKA	Q8tu24 methanopyru
14	231.5	20.0	283	1 DAPA_METH	Q26892 methanobact
15	226	19.5	291	1 DAPA_METMA	Q8px17 methanosarc
16	225.5	19.5	290	1 NANA_FUSNN	Q8rdn6 fusobacteri
17	224.5	19.4	291	1 DAPA_NEIMA	Q9juu9 neisseria m
18	222.5	19.2	301	1 DAPA_CORGL	P19808 corynebacte
19	222	19.2	294	1 DAPA_AQUA3	O67216 aquifex aeo
20	220.5	19.0	287	1 DAPA_THETN	Q8bi5 thermoaer
21	218.5	18.9	288	1 NANA_CLOPE	Q384k9 clostridium
22	218.5	18.9	301	1 YGHH_ECOLI	P39359 escherichia
23	218	18.8	309	1 YAGE_ECOLI	P75682 escherichia
24	217.5	18.8	291	1 DAPA_NEIMB	Q9jzr4 neisseria m
25	216.5	18.7	291	1 DAPA_METAC	Q8thp1 methanosarc
26	216	18.6	302	1 DAPA_PROMA	P49423 prochloroco
27	213.5	18.4	301	1 DAPA_COREF	Q8rqm8 corynebacte
28	213	18.4	295	1 DAPA_OCRH	Q8eqj1 oceanobacil
29	208	17.9	294	1 DAPA_ANASP	Q9qyvl anabaena sp
30	207.5	17.9	294	1 DAPA_STAEP	Q8cp96 staphylococ
31	206	17.8	291	1 DAPA_CLOPE	Q8xj56 clostridium
32	206	17.8	307	1 DAPA_LEPIN	Q8f132 leptospira
33	205.5	17.7	294	1 DAPA_AGRTS	Q8ug13 agrobacteri

34	204.5	17.6	292	1 DAP2_BACHD	Q9ka91 bacillus ba
35	203.5	17.6	290	1 DAP1_BACSU	Q04796 bacillus su
36	202.5	17.5	294	1 DAP1_RHIME	Q92x55 rhizobium m
37	202.5	17.5	294	1 DAPA_BALSO	Q8y099 raietonia s
38	202	17.4	296	1 DAPA_BRAJA	Q9rh76 bradyrhizob
39	201.5	17.4	289	1 DAPA_METJA	Q57695 methanococc
40	201	17.3	301	1 DAPA_SVNY3	Q55513 syntecocyst
41	200	17.3	298	1 DAPA_HAEIN	P43797 haemophilus
42	199.5	17.2	301	1 DAPA_BIFLO	Q8g527 bifidobacte
43	199	17.2	296	1 DAPA_SYNEL	Q8djk4 syntecococc
44	197.5	17.0	305	1 DAP2_STRCO	Q9x9w0 streptomyce
45	197.5	17.0	333	1 MOSA_RHIME	Q07607 rhizobium m

ALIGNMENTS

RESULT 1	ID	NANA_ECOLI	STANDARD;	PRT;	296 AA.
AC	P06995;				
DT	01-APR-1988	(Rel. 07, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	N-acetylneuraminase	lyase (EC 4.1.3.3) (N-acetylneuraminic acid			
DE	aldolase)	(N-acetylneuraminase pyruvate-lyase) (Sialic acid lyase)			
DE	(Sialate lyase)	(Sialic acid aldolase) (NALase).			
GN	NANA OR NPL OR B3225 OR Z4583 OR ECS4098 OR SF3261 OR S3478.				
OS	Escherichia coli,				
OS	Shigella flexneri,				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID=562, 83334, 623;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	SPECIES=E.coli; STRAIN=K12;				
RX	MEDLINE=86093682; PubMed=3909108;				
RA	Onta Y., Watanabe K., Kimura A.;				
RT	"Complete nucleotide sequence of the E. coli N-acetylneuraminase				
RT	lyase."				
RL	Nucleic Acids Res. 13:8843-8852 (1985).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=E.coli; STRAIN=JEL1011;				
RA	Kawakami B., Kudo T., Narahashi Y., Horikoshi K.;				
RT	"Nucleotide sequence of the N-acetylneuraminase lyase gene of				
RT	Escherichia coli."				
RL	Agric. Biol. Chem. 50:2155-2158 (1986).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=E.coli; STRAIN=K12 / MG1655;				
RX	MEDLINE=97426617; PubMed=9278503;				
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,				
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,				
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,				
RA	Mau B., Shao Y.;				
RT	"The complete genome sequence of Escherichia coli K-12."				
RL	Science 277:1453-1474 (1997).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=E.coli; STRAIN=0157:H7 / EDL933 / ATCC 700927;				
RX	MEDLINE=21074935; PubMed=11206551;				
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,				
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,				
RA	Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,				
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,				
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,				
RA	Weich R.A., Blattner F.R.;				
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."				
RN	Nature 409:529-533 (2001).				
RN	[5]				
RP	SEQUENCE FROM N.A.				

CC SPECIES=E.coli; STRAIN=0157:H7 / RIMD 0509952;
RA MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka Y., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara K., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RT DNA Res. 8:11-22(2001).
RL [6]
RN SEQUENCE FROM N.A.
RP SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Heu Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [7]
RN SEQUENCE FROM N.A.
RP SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan N.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perra N.T., Payne S.M., Ruyten-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
RN [8]
RN SEQUENCE OF 1-23, AND CHARACTERIZATION.
RP SPECIES=E.coli; STRAIN=K12 / C600 / SF8;
RX MEDLINE=91264813; PubMed=1646603;
RA Aisaka K., Igarashi A., Yamaguchi K., Uwajima T.;
RT "Purification, crystallization and characterization of
RT N-acetylneuraminase lyase from Escherichia coli.";
RL Biochem. J. 276:541-546(1991).
RN [9]
RN X-RAY CRYSTALLOGRAPHY {2.2 ANGSTROMS}.
RP SPECIES=E.coli;
RX MEDLINE=94363237; PubMed=8081752;
RA Izard T., Lawrence M.C., Malby R.L., Lilley G.G., Colman P.M.;
RT "The three-dimensional structure of N-acetylneuraminase lyase from
RT Escherichia coli.";
RL Structure 2:361-369(1994).
RN [-0]
RN X-RAY CRYSTALLOGRAPHY {2.45 ANGSTROMS} OF COMPLEXES WITH
HYDROXYPYRUVATE AND PYRUVATE.
RP SPECIES=E.coli;
RX MEDLINE=97199395; PubMed=9047371;
RA Lawrence M.C., Barbosa J.A., Smith B.J., Hall N.E., Pilling P.A.,
RA Ooi H.C., Marcuccio S.M.;
RT "Structure and mechanism of a sub-family of enzymes related to N-
RT acetylneuraminase lyase.";
RL J. Mol. Biol. 266:381-399(1997).
CC [-] FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
CC (sialic acid) to form pyruvate and N-acetylmannosamine via a
CC Schiff base intermediate.
CC [-] CATALYTIC ACTIVITY: N-acetylneuraminase = N-acetyl-D-mannosamine +
CC pyruvate.
CC [-] ENZYME REGULATION: Inhibited by reduction with NaBH(4), and by
CC Cu(2+) ions, p-chloromercuribenzoate and N-bromosuccinimide.
CC [-] PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.
CC [-] SUBUNIT: Homotetramer.
CC [-] SUBCELLULAR LOCATION: Cytoplasmic.
CC [-] INDUCTION: By N-acetylneuraminase.
CC [-] MISCELLANEOUS: Optimum temperature for activity is 80 degrees
CC Celsius and optimum pH is 6.5-7.0.
CC [-] SIMILARITY: Belongs to the DHPS family. Nana subfamily.
RN -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
RN -----
DR EMEL; X03345; CAA27051.1; -;
DR EMEL; D00067; BAA00046.1; -;
DR EMEL; U18997; AAA58027.1; -;
DR EMEL; AE000402; AAC76257.1; -;
DR EMEL; AE000550; AAG59353.1; -;
DR EMEL; AF002564; BAB37521.1; -;
DR EMEL; AE015337; AAB44725.1; -;
DR EMEL; AE016989; AAB18538.1; -;
DR PIR; B91141; B91141;
DR PIR; E85986; E85986;
DR PIR; JP0002; WZECN;
DR PDB; 1NAL; 15-SEP-95.
DR PDB; 1FDY; 22-OCT-97.
DR PDB; 1FDZ; 22-OCT-97.
DR ECO2DBASE; G032.6; 6TH EDITION.
DR EcoGene; EG10637; nana.
DR HAMAP; MF_01237; -; 1.
DR InterPro; IPR002220; DHDS.
DR InterPro; IPR005264; Nana.
DR Pfam; PF00701; DHDS; 1.
DR PRINTS; PR00146; DHPICNTHASE.
DR ProDom; PD001859; DHDS; 1.
DR TIGRfams; TIGR00683; nana; 1.
DR PROSITE; PS00665; DHDS_1; 1.
DR PROSITE; PS00666; DHDS_2; 1.
DR Carbohydrate metabolism; Lyase; Schiff base; 3D-structure;
KW Complete proteome.
FT INIT MET 0
FT ACT_SITE 136 136 INVOLVED IN PROTON TRANSFER DURING
FT ACT_SITE 164 164 CLEAVAGE.
FT ACT_SITE 164 164 SCHIFF-BASE WITH N-ACETYLNEURAMINATE.
FT CONFLICT 69 69 A -> G (IN REF. 1).
FT CONFLICT 83 83 S -> T (IN REF. 1).
FT CONFLICT 281 281 L -> Q (IN REF. 1).
FT STRAND 7 10
FT STRAND 15 15
FT TURN 17 18
FT STRAND 21 21
FT STRAND 23 35
FT TURN 36 37
FT STRAND 40 43
FT TURN 46 49
FT HELIX 50 52
FT HELIX 55 69
FT TURN 70 72
FT STRAND 74 78
FT HELIX 84 96
FT TURN 97 98
FT STRAND 101 105
FT HELIX 114 127
FT TURN 128 130
FT STRAND 133 137
FT HELIX 139 142
FT HELIX 148 154
FT TURN 155 156
FT TURN 158 159
FT STRAND 160 165
FT HELIX 170 179
FT TURN 181 182
FT STRAND 184 187
FT HELIX 190 192
FT TURN 193 198
FT TURN 199 200
FT STRAND 203 206
FT TURN 207 208

TIGRPAWS: TIGR00683; nana; 1.
 DR PROSITE; PS00665; DHPS 1; 1.
 DR PROSITE; PS00666; DHPS 2; 1.
 KW Carbohydrate metabolism; lyase; Schiff base; Complete proteome.
 FT INIT MET 0
 FT ACT_SITE 136 136 INVOLVED IN PROTON TRANSFER DURING
 FT ACT_SITE 164 164 CLEAVAGE (BY SIMILARITY).
 FT ACT_SITE 164 164 SCHIFF-BASE WITH N-ACETYLNEURAMINATE
 FT ACT_SITE 164 164 (BY SIMILARITY).
 SQ SEQUENCE 296 AA; 32463 MW; DDDCEB3DA9FC601 CRC64;
 Query Match 23.6%; Score 273.5; DB 1; Length 296;
 Best Local Similarity 34.3%; Pred. No. 9.3e-15;
 Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;
 QY 8 LOGVAATITPMTEINGEINFSVIGQYVDYLKVGKQKNIKFNVTGEGLSLSVSRQ-- 65
 DB 4 LRGVMAALLTFDQOQALDKASIRLVQFNI--QQGIDGLYVGGSTGSAFVQSLSERQVL 62
 QY 66 --VAEHWTKGKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFELKPTWKDI 123
 DB 63 EIVAE--AKGKIL---IAHVCVSTAESQQLAASAKRYGFDVAVSATEFY-PPSPEE 116
 QY 124 LINFKEVAAAPALPFYHYHIPALTGKIRAEELLDGILDKIPTFQGLKFSDTLLDFG 183
 DB 117 HCDHYRAIIDSADGLPMVYVNIIPALSGVKLTLDQI--DTLVTLPGVGALKQTSGLDQYME 174
 QY 184 QCVDNQOQFAFLF-GVDEQLLSALVNGTGAAGS 218
 DB 175 QI---RREHPDLVLYNGYDEIFASGLLAGADGGIGS 207
 RESULT 3
 NANA_SALTI
 ID NANA_SALTI STANDARD; PRT: 297 AA.
 AC Q823F0;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE N-acetylneuraminatase (EC 4.1.3.3) (N-acetylneuraminic acid
 aldolase) (N-acetylneuraminatase pyruvate-lyase) (Sialic acid lyase)
 DE (Sialate lyase) (Sialic acid aldolase).
 GN NANA OR STY3520 OR T3256.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Farhrell J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogan A., Larsen T.S., Leather S., McAlle S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
 (sialic acid) to form pyruvate and N-acetylmannosamine via a

Query Match 23.7%; Score 274.5; DB 1; Length 296;
 Best Local Similarity 34.3%; Pred. No. 7.7e-15;
 Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;
 QY 8 LOGVAATITPMTEINGEINFSVIGQYVDYLKVGKQKNIKFNVTGEGLSLSVSRQ-- 65
 DB 4 LRGVMAALLTFDQOQALDKASIRLVQFNI--QQGIDGLYVGGSTGSAFVQSLSERQVL 62
 QY 66 --VAEHWTKGKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFELKPTWKDI 123
 DB 63 EIVAE--AKGKIL---IAHVCVSTAESQQLAASAKRYGFDVAVSATEFY-PPSPEE 116
 QY 124 LINFKEVAAAPALPFYHYHIPALTGKIRAEELLDGILDKIPTFQGLKFSDTLLDFG 183
 DB 117 HCDHYRAIIDSADGLPMVYVNIIPALSGVKLTLDQI--NTLVTLPGVGALKQTSGLDQYME 174
 QY 184 QCVDNQOQFAFLF-GVDEQLLSALVNGTGAAGS 218
 DB 175 QI---RREHPDLVLYNGYDEIFASGLLAGADGGIGS 207
 RESULT 2
 NANA_ECOL6
 ID NANA_ECOL6 STANDARD; PRT: 296 AA.
 AC Q8FD58;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE N-acetylneuraminatase (EC 4.1.3.3) (N-acetylneuraminic acid
 aldolase) (N-acetylneuraminatase pyruvate-lyase) (Sialic acid lyase)
 DE (Sialate lyase) (Sialic acid aldolase).
 GN NANA OR NANA OR C3979.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=2.7992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6.H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22398234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P., Stroud D.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
 (sialic acid) to form pyruvate and N-acetylmannosamine via a
 Schiff base intermediate (By similarity).
 CC -!- CATALYTIC ACTIVITY: N-acetylneuraminatase = N-acetyl-D-mannosamine +
 pyruvate.
 CC -!- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.
 CC -!- SUBUNIT: Homotrimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the DHPS family. Nana subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annunce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; A016767; AAN82419.1; -;
 DR HAMAP; MF 01237; -; 1.
 DR InterPro; IPR002220; DHPS.
 DR InterPro; IPR005264; Nana.
 DR Pfam; PF00701; DHPS; 1.
 DR PRINTS; PR00146; DHPCSNTHASE.
 DR ProDom; PD001959; DHPS; 1.

```

CC Schiff base intermediate (By similarity).
CC -!- CATALYTIC ACTIVITY: N-acetylneuraminase = N-acetyl-D-mannosamine +
CC pyruvate.
CC -!- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the DHPS family. NaaA subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AL627278; CAF07856.1; -
CC EMBL: AEC16845; AAC07091.1; -
CC HAMAP: MF_01237; -; 1
CC InterPro: IPR002220; DHPS.
CC InterPro: IPR005284; NaaA.
CC Pfam: PF02701; DHPS; 1.
CC PRINTS: PR0146; DHPICNTHASE.
CC ProDom: PD01859; DHPS; 1.
CC TIGRfam: TIGR00683; naaA; 1.
CC PROSITE: PS00665; DHPS 1; 1.
CC PROSITE: PS00666; DHPS_2; 1.
CC K0 Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.
FT ACT_SITE 137 137 INVOLVED IN PROTON TRANSFER DURING
FT FT CLEAVAGE (BY SIMILARITY).
FT FT SCHAFF-BASE WITH N-ACETYLNEURAMINATE
FT FT (BY SIMILARITY).
FT ACT_SITE 165 165
FT ACT_SITE 165 165
FT SEQUENCE 297 AA; 32482 MW; D8FEB0F1C4895546 CRC64;
CC -----
Query Match 23.5%; Score 272.5; DB 1; Length 237;
Best Local Similarity 31.9%; Pred. No. 1.1e-14;
Matches 68; Conservative 43; Mismatches 95; Indels 7; Gaps 5;
CC -----
QY 6 KX-LQGLVAATITPWTENGENSEFVIGYVDLVKEQGVKNIFVNGTGTGGLSLSVSERRQ 65
DB 3 KALQGVAAALLTFDHCQQQLDSSLSRLVAFNI-QGIDGLGYGSGTGEAFVQSLAEREQ 61
QY 66 VAEWYTKGDKLDQWYIIVGALSLLKESQELAQHMAEIGADGIAVIAPFLKPKWTKDILI 125
DB 62 VLEIVAEAAAGKI-TLTAHVGTGVSTAESQQLAGAAKRYGFDASVATPFYF-PFSFEHC 119
QY 126 NEFLKEVAAPALPFYTHIPALTGVKIRAEILLDGLDKPIETFGQLKFSDTLLDFGQC 185
DB 120 DHYRAILDSADGLPMWYINYPALSGVKLTLDQI--NLVTLPGVNAKQISGDLFQMEQI 177
QY 186 VDQNRQQCAFALFGVDQQLLSALVMGAIGAWGS 218
DB 178 --RRAPDLVLVXGYDEIFAASGLIAGADGGTGS 208
CC -----
RESULT 4
NANA SALTY STANDARD; PRT; 297 AA.
ID ID NANA SALTY STANDARD; PRT; 297 AA.
AC AC Q6ZLQ6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE N-acetylneuraminase lyase (EC 4.1.3.3) [N-acetylneuraminic acid
DE aldolase] (N-acetylneuraminase pyruvate-lyase, (Sialic acid lyase)
DE (Sialate lyase) (Sialic acid aldolase).
GN NANA OR STM3339.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
CC NCBI_TaxID=602;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;

```

N-acetylneuraminase lyase (EC 4.1.3.3) (N-acetylneuraminic acid aldolase) (N-acetylneuraminase pyruvate-lyase) (Sialic acid lyase) (Sialate lyase) (Sialic acid aldolase).

NANA OR PM1715.

OS Pasteurella multocida.

CC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

CC Pasteurellaceae; Pasteurella.

OX NCBI_TaxID=747;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Pm70;

RX MEDLINE=21145856; PubMed=11248100;

RA May B.J., Zhang Q., Li L.B., Paustian V.L., Whittam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida Pm70.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid (sialic acid) to form pyruvate and N-acetylmannosamine via a Schiff base intermediate (By similarity).

CC -!- CATALYTIC ACTIVITY: N-acetylneuraminase = N-acetyl-D-mannosamine + pyruvate.

CC -!- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.

CC -!- SUBUNIT: Homotrimer (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: Belongs to the DHPS family. Nana subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AE006207; AAK03799.1; -.

DR HAMAP: MF_01237; -; 1.

DR InterPro: IPR002220; DHPS.

DR InterPro: IPR005264; Nana.

DR Pfam: PF00701; DHPS; 1.

DR PRINTS: PR00146; DHPICNTHASE.

DR PRODOM: PD001859; DHPS; 1.

DR TIGRFAMs: TIGR00683; nana; 1.

DR PROSITE: PS00665; DHPS_1; 1.

DR Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.

FT ACT_SITE 136 136 INVOLVED IN PROTON TRANSFER DURING CLEAVAGE (BY SIMILARITY).

FT ACT_SITE 164 164 SCHIFF-BASE WITH N-ACETYLNEURAMINATE (BY SIMILARITY).

FT ACT_SITE 164 164

FT ACT_SITE 293 AA; 32607 MW; 69AE7B7729340545 CRC64;

SEQ SEQUENCE 293 AA; 32607 MW; 69AE7B7729340545 CRC64;

Query Match 21.4%; Score 247.5; DB 1; Length 293;

Best Local Similarity 29.8%; Pred. No. 1.1e-12;

Matches 68; Conservative 41; Mismatches 82; Indels 37; Gaps 6;

QY 6 KKQGLVAATTPMTENGESVIGQYVDYLNKEQGVKNIFVNGTTEGGLSLVSERRQ 65

DB 2 KNEKGFSAIVSNFADGSENEKGLRQIVRYNDKMKVDGLYGVGSGENFMFLSTEEKE 61

QY 66 VAEWVTGKDKLDQVLIHVGNLSKESQELAQHAABIGADGIAVIAPFFLKPWTKDILI 125

DB 62 IFRANDEAKDEI-ALIAQGVSNLQELAEIKGKATELGYDLSAVTFYYK----- 112

QY 126 NFLKEVAAAAPALPFYY-----YHIPALGVKIRAEELDGLDKIPTQGL 172

DB 113 -----FSFPEIKHYVDSIIIEATGNMIVYSIPFLGVNIGVQF--GELYKNPKVLGV 163

QY 173 KFSSTD--LLDFGQVDONQQQFAFLFGVDQELLSALWAGTAVGS 218

DB 164 KFTAGDFILLERUKKAYPNH-----LIWAGFDEMMLPAAS-GVDGAIGS 207

RESULT 6

NAA2_EC0L6

ID NAA2_EC0L6 STANDARD; PRT; 305 AA.

AC Q8FDD7;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DE N-acetylneuraminase lyase 2 (EC 4.1.3.3) (N-acetylneuraminic acid aldolase 2) (N-acetylneuraminase pyruvate-lyase 2) (Sialic acid lyase 2) (Sialate lyase 2) (Sialic acid aldolase 2).

GN NANA2 OR C3639.

OS Escherichia coli O6.

CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

CC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=217992;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O6:H1 / CFT373 / ATCC 700928;

RX MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,

RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,

RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;

RT "Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).

CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid (sialic acid) to form pyruvate and N-acetylmannosamine via a Schiff base intermediate (By similarity).

CC -!- CATALYTIC ACTIVITY: N-acetylneuraminase = N-acetyl-D-mannosamine + pyruvate.

CC -!- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.

CC -!- SUBUNIT: Homotrimer (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: Belongs to the DHPS family. Nana subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AE016766; AAN82087.1; -.

DR HAMAP: MF_01237; -; 1.

DR InterPro: IPR002220; DHPS.

DR InterPro: IPR005264; Nana.

DR Pfam: PF00701; DHPS; 1.

DR PRINTS: PR00146; DHPICNTHASE.

DR PRODOM: PD001859; DHPS; 1.

DR TIGRFAMs: TIGR00683; nana; 1.

DR PROSITE: PS00665; DHPS_1; 1.

DR Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.

FT ACT_SITE 137 137 INVOLVED IN PROTON TRANSFER DURING CLEAVAGE (BY SIMILARITY).

FT ACT_SITE 165 165 SCHIFF-BASE WITH N-ACETYLNEURAMINATE (BY SIMILARITY).

FT ACT_SITE 305 AA; 34098 MW; 50482F39B3C701A3 CRC64;

SEQ SEQUENCE 305 AA; 34098 MW; 50482F39B3C701A3 CRC64;

Query Match 21.3%; Score 246.5; DB 1; Length 305;

Best Local Similarity 28.8%; Pred. No. 1.4e-12;

Matches 61; Conservative 52; Mismatches 92; Indels 7; Gaps 5;

QY 7 KQGLVAATTPMTENGESVIGQYVDYLNKEQGVKNIFVNGTTEGGLSLVSERRQ 66

DB 4 EFKGVSALPTPYDQSQIDMESLRKLIIRFNI-EQNIKGLYGVGSGTGEAFIQNAEREKI 62

QY 67 AEWVTGKDKLDQVLIHVGNLSKESQELAQHAABIGADGIAVIAPFFLKPWTKDILI 126

DB 63 LETVADESQRL-TLIAHVGIGISTAESEVLAQAQKYGYTHAISAVTFYFFFSFEHCII 121

QY 127 FLKEVAAAAPALPFYYVHIIPALGVKIRAEELDGLDKIPTQGLKFSDDLLDFGQCV 186

DB 122 Y-RKIIDSAGLPMVYVYNIPALSGVRFSLDQINE--LVTIPREVCALKQTSGLFQWQI- 177
 QY 187 QNRQOQAFILFGVDEQLLSALVMGATGAVGS 218
 DB 178 -KRNHPELVLYNGYDEIFASGLIAGADGGIGS 208

RESULT 7

NANA STAAW
 ID NANA STAAW STANDARD; PRT; 293 AA.
 AC Q99WR1.
 DT 13-OCT-2003 (Rel. 42, Created)
 DT 13-OCT-2003 (Rel. 42, Last sequence update)
 DT 19-OCT-2003 (Rel. 42, Last annotation update)
 DE N-acetylneuraminatase lyase (EC 4.1.3.3) (N-acetylneuraminic acid
 aldolase) (N-acetylneuraminatase pyruvate-lyase) (Sialic acid lyase)
 DE (Sialate lyase) (Sialic acid aldolase).
 GN NANA OR SAV0315 OR SA0304.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OS Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=158878, 158879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mu50 / ATCC 700699, and N315;
 RX MEDLINE=21311952; PubMed=1418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Kuroda M., Ohta T., Uchiyama I., Nagai Y., Igarashi H., Ito T.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Igarashi H., Hosoyama A.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
 RA Kanesaka M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus";
 RL Lancet 357:1225-1240(2001).
 CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
 (Sialic acid) to form pyruvate and N-acetylmannosamine via a
 Schiff base intermediate (By similarity).
 CC -!- CATALYTIC ACTIVITY: N-acetylneuraminatase = N-acetyl-D-mannosamine +
 pyruvate.
 CC -!- PATHWAY: N-acetylneuraminic acid (Sialic acid) utilization.
 CC -!- SUBUNIT: Homotetramer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the DHPS family. Nana subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AP003359; BAB56477.1; -;
 CC EMBL; AP003130; BAB4528.1; -;
 CC F01; E89796; E89796.
 CC HSP; P06995; INAL.
 CC SWISS-2DPAGE; Q99WR1; STAAW.
 CC HAMAP; MF 01237; -; 1.
 CC InterPro; IPR002220; DHPS.
 CC Pfam; PF00701; DHPS; 1.
 CC PRINTS; PR00146; DHPICNTASE.
 CC PROSITE; PD001859; DHPS; 1.
 CC PROSITE; PS00665; DHPS; 1; FALSE NEG.
 CC PROSITE; PS00666; DHPS; 2; FALSE NEG.
 KM Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.
 FT ACT_SITE 137 137 INVOLVED IN PROTON TRANSFER DURING
 FT ACT_SITE 165 165 CLEAVAGE (BY SIMILARITY).
 FT ACT_SITE 165 165 SCHIFF-BASE WITH N-ACETYLNEURAMINATE
 (BY SIMILARITY).
 FT SEQUENCE 293 AA; 33100 MW; 4D07DEFFDCLAF701 CRC64;

Query Match 20.9%; Score 242.5; DB 1; Length 293;
 Best Local Similarity 30.0%; Pred. No. 2.7e-12;
 Matches 68; Conservative 46; Mismatches 78; Indels 35; Gaps 8;
 QY 6 KKLGLVAATITETENGSEINFSVIGQYVDYLAVKEQGVKNIEVNGTTGGLSISVSRPQ 65
 DB 3 KDLKGLYALIVFDENGQVNEQGLKQIAQNAISTEELDLTYNGSSGENFLINTEQKQ 62
 QY 66 ---VAEEWVTGKMDQVVIHVGLSLKESQELAQAHAIGADGLAVIAPFFLKPTWKD 122
 DB 63 VFKVAKEAV---GDKV-KLIAQVGSLSLNEALELCKYATELCYDALSAVTPEY-PEETE 117
 QY 123 ILINELKEVARAAPALPFYHYHIPALTGVKIRAE-----ELDLGILDKITFP---Q3 171
 DB 118 EIRDFYDIIIEATQN-NMIIYAIPLDTGVNISIEQFSELFNEHKIVGVKYTAPNFFLLER 176
 QY 172 LKFSPTDILLDFGQVDQNRQOQAFILFGVDEQLLSALVMGATGAVGS 218
 DB 177 IRKAFPDKL-----ILSGFDEMLVQATISGVGAGS 208

RESULT 8

NANA STAAW
 ID NANA STAAW STANDARD; PRT; 293 AA.
 AC Q99WR1.
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE N-acetylneuraminatase lyase (EC 4.1.3.3) (N-acetylneuraminic acid
 aldolase) (N-acetylneuraminatase pyruvate-lyase) (Sialic acid lyase)
 DE (Sialate lyase) (Sialic acid aldolase).
 GN NANA OR MW0292.
 OS Staphylococcus aureus (strain MW2).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=196620;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22040717; PubMed=12044378;
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
 RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
 RA Yamamoto K., Hiramatsu K.;
 RT "Genome and virulence determinants of high virulence community-
 acquired MRSA";
 RL Lancet 359:1819-1827(2002).
 CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
 (Sialic acid) to form pyruvate and N-acetylmannosamine via a
 Schiff base intermediate (By similarity).
 CC -!- CATALYTIC ACTIVITY: N-acetylneuraminatase = N-acetyl-D-mannosamine +
 pyruvate.
 CC -!- PATHWAY: N-acetylneuraminic acid (Sialic acid) utilization.
 CC -!- SUBUNIT: Homotetramer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the DHPS family. Nana subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AP004823; BAB94157.1; -;
 CC HAMAP; MF 01237; -; 1.
 CC InterPro; IPR002220; DHPS.
 CC Pfam; PF00701; DHPS; 1.
 CC PRINTS; PR00146; DHPICNTASE.
 CC PROSITE; PD001859; DHPS; 1.
 CC PROSITE; PS00665; DHPS; 1; FALSE NEG.
 CC PROSITE; PS00666; DHPS; 2; FALSE NEG.
 KM Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.
 FT ACT_SITE 137 137 INVOLVED IN PROTON TRANSFER DURING

FT ACT_SITE 165 165 CLEAVAGE (BY SIMILARITY).
 FT SCHIFF-BASE WITH N-ACETYLNEURAMINATE
 FT (BY SIMILARITY).
 SQ SEQUENCE 293 AA; 33042 MW; 62D7DEFFCFA431 CRC64;
 Query Match 20.9%; Score 242.5; DB 1; Length 293;
 Best Local Similarity 30.0%; Pred. No. 2.7e-12;
 Matches 66; Conservative 46; Mismatches 78; Indels 35; Gaps 8;
 QY 6 KKLGLVAATITPMTEINGENFVSIGVYDYLKQGVKQVFNVTGTGEGSLSVSRERQ 65
 DB 3 KDLKGLYARLLVPDENGQVNEQLKQIAQNAITEELDGLYNGSSGFLLNTEQKKQ 62
 QY 66 ---VAEHWYTKGDKLDQVIVHVGALSLSKESQELAAHAEIGADGIAVIAPFFKLPWKD 122
 DB 63 VFKVAKAV---GDKV-KLIAQVGLSLDNEALGKYTELGYSALSAVTPFYV-PTFE 117
 QY 123 ILINFLKEVAAPALPFYVYHIIHVGALSLSKESQELAAHAEIGADGIAVIAPFFKLPWKD 171
 DB 118 EIKDYFETIEATQN-NMIIYALFDLGVNISIQFSELFNHEKIVGVKTAPEAFLLER 176
 QY 172 LKFSDDTLDFGQCVDQNRQQQAFLEFGVDEQLLSALVMGATGAVGS 218
 DB 177 IRKAFPKL-----ILSGFDEMLVQATISGVGGAIGS 208
 RESULT 9
 NANA LACPL STANDARD; PRT; 292 AA.
 ID NANA LACPL
 AC P59407;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE N-acetylneuraminatase (EC 4.1.3.3) (N-acetylneuraminic acid
 aldolase) (N-acetylneuraminatase pyruvate-lyase) (Sialic acid lyase)
 DE (Sialate lyase) (Sialic acid aldolase).
 GN NANA OR LP 3568.
 OS Lactobacillus plantarum.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1590;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCIMB 8826 / WCPS1;
 RX MEDLINE=22480296; PubMed=12566566;
 RA Kleerebezem Y., Boekhorst J., van Xranenburg R., Molenaar D.,
 RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink R.V.,
 RA Fiers W.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
 RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
 RA De Vos W.M., Siezen R.J.;
 RT "Complete genome sequence of Lactobacillus plantarum WCPS1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
 CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
 (sialic acid) to form pyruvate and N-acetylmannosamine via a
 Schiff base intermediate (By similarity).
 CC -!- CATALYTIC ACTIVITY: N-acetylneuraminatase = N-acetyl-D-mannosamine +
 Pyruvate.
 CC -!- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.
 CC -!- SUBUNIT: Homotrimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the DHDPS family. Nana subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; ALG95262; CA965647.1; -.
 CC HAMAP; MF 01237; -; 1.
 CC InterPro; IPR002220; DHDPS.

DR PFam: PF00701; DHDPS: 1.
 DR PRINTS: PR00146; DHDPS: 1.
 DR PROSITE: PS00665; DHDPS: 1; 1.
 DR PROSITE: PS00666; DHDPS: 2; FALSE NEG.
 KW Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.
 FT ACT_SITE 135 135 INVOLVED IN PROTON TRANSFER DURING
 FT CLEAVAGE (BY SIMILARITY).
 FT ACT_SITE 163 163 SCHIFF-BASE WITH N-ACETYLNEURAMINATE
 FT (BY SIMILARITY).
 SQ SEQUENCE 292 AA; 32708 MW; 4184967AE9C91747 CRC64;
 Query Match 20.8%; Score 241.5; DB 1; Length 292;
 Best Local Similarity 31.9%; Pred. No. 3.3e-12;
 Matches 69; Conservative 38; Mismatches 94; Indels 15; Gaps 6;
 QY 6 KKLGLVAATITPMTEINGENFVSIGVYDYLKQGVKQVFNVTGTGEGSLSVSRERQ 65
 DB 3 KKL--LYAAQMTAFDKDNINLDGIRALVRNIDVKNVGLYVCGSTGEAFMLNTEKKQ 60
 QY 66 VAEHWYTKGDKLDQVIVHVGALSLSKESQELAAHAEIGADGIAVIAPFFKLPWKD 122
 DB 61 VMTVYDEANGAID-LVAQVGLSLNKEAKELAKATDGLYPKLSAVTFYNYNFTPEQKD 119
 QY 123 ILINFLKEVAAPALPFYVYHIIHVGALSLSKESQELAAHAEIGADGIAVIAPFFKLPWKD 182
 DB 120 YNEILKDVON-----KLLIYIPALTGVALTTPQFAB--LFENPKIIGIXYNADFVLL 172
 QY 183 GQCVDQNRQQQAFLEFGVDEQLLSALVMGATGAVGS 218
 DB 173 ERV--RNAFPDKLLISGFDEMLLPALALNVGDCIGS 206
 RESULT 10
 DAPA HALK1 STANDARD; PRT; 304 AA.
 ID DAPA HALK1
 AC QHSH19;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dihydrodipicolinate synthase (EC 4.2.1.52) (DHDPS).
 GN DAPA OR VNG0444G.
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20504483; PubMed=11016950;
 RX NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Laaky S.R., Balliga N.S., Thorsson V., Sirogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
 RA Leithausen B., Keller K., Cruz R., Banson M.J., Hough D.W.,
 RA Madocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlshocher M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebbardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
 RT "Genome sequence of Halobacterium species NRC-1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 CC -!- CATALYTIC ACTIVITY: L-aspartate 4-semialdehyde + pyruvate =
 CC dihydrodipicolinate + 2 H(2O).
 CC -!- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
 CC semialdehyde; first step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the DHDPS family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; ALG95262; CA965647.1; -.
 CC HAMAP; MF 01237; -; 1.
 CC InterPro; IPR002220; DHDPS.


```

DE aldolase) (N-acetylneuraminase pyruvate-lyase) (Sialic acid lyase)
DE (Sialate lyase) (Sialic acid aldolase).
GN NANA OR RBO289 OR SMB20299.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OS Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=1021;
RC MEDLINE=21396508; PubMed=1481431;
RX Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vortelselt F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-
RL fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
CC (sialic acid) to form pyruvate and N-acetylmannosamine via a
CC Schiff base intermediate (By similarity).
CC -!- CATALYTIC ACTIVITY: N-acetylneuraminate + N-acetyl-D-mannosamine +
CC pyruvate.
CC -!- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the DHDPS family. Nana subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL603642; CAC48689.1; -
DR PIR; A95878; A95878.
DR HAMAP; MF 01237; -; 1.
DR InterPro; IPR002220; DHDPS.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHPICNTHASE.
DR PRODOM; PD001859; DHDPS; 1.
DR PROSITE; PS00665; DHDPS_1; FALSE NEG.
DR PROSITE; PS00666; DHDPS_2; FALSE NEG.
KW Carbohydrate metabolism; Lyase; Schiff base; Plasmid;
KW Complete proteome.
FT ACT_SITE 134 134 INVOLVED IN PROTON TRANSFER DURING
FT ACT_SITE 161 161 CLEAVAGE (BY SIMILARITY).
FT ACT_SITE 161 161 SCHIFF-BASE WITH N-ACETYLNEURAMINATE
FT (BY SIMILARITY).
SQ SEQUENCE 299 AA; 32380 MW; 0FA610AD53886851 CRC64;
Query Match 20.5%; Score 237.5; DB 1; Length 299;
Best Local Similarity 29.2%; Pred. No. 7e-12;
Matches 62; Conservative 49; Mismatches 92; Indels 9; Gaps 7;
Qy 7 KLOGLVAATTPMTENGESVITGVYDVKVKEQGVKNIFVNGTTGEGLSLSVSRQW 66
Db 2 KLEGIYSALLTFPESEDSIDRQALGALVDEQVR-LGIDGVYVGGSGEAMLCQLDERADY 60
Qy 67 ABEWVTGKDKLDQVIVHVGALSKEQELAQHAAEIGADGIAVIAPFLKPTWKDILIN 126
Db 61 LSDVAAASAGRL-TLIAFVGTIARDLRLSCHAAKSGYQALSIPIPPYDFSRPEVNAH 119
Qy 127 FLKEVAAAAPALPFYYHIPALTGVKIRAEELLDGILDKIPTFQGLKFSDDTLDFGQCV 186
Db 120 Y-RELADVS-ALPLIVNFPARTS-GFTLPPELVE--LLSHNPI-GIKHTSSDMQLERI- 173
Qy 187 DONFQQQFAFLFGVDEQLLSALVMGATGAVS 218
Db 174 -RHAVPDAIVNGYDEMCLAGFAMGAQGAIGT 204

```

```

RESULT 13
DAPA METKA STANDARD; PRT; 300 AA.
ID DAPA METKA
AC QRTUZA;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydropicolinate synthase (EC 4.2.1.52) (DHDPS).
GN DAPA OR MKI607.
OS Methanopyrus kandleri.
OS Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OX Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=AV19 / DSM 6324 / JCM 9639;
RC MEDLINE=21927647; PubMed=11930014;
RX Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Koonin E.V., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlyavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -!- CATALYTIC ACTIVITY: L-aspartate 4-semialdehyde + pyruvate =
CC dihydropicolinate + 2 H(2)O.
CC -!- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
CC semialdehyde; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the DHDPS family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE010450; AAM02823.1; -
DR HAMAP; MF 00418; -; 1.
DR InterPro; IPR005263; Dapa.
DR InterPro; IPR002220; DHDPS.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHPICNTHASE.
DR PRODOM; PD001859; DHDPS; 1.
DR TIGRFAMs; TIGR00674; dapa; 1.
DR PROSITE; PS00665; DHDPS_1; 1.
DR PROSITE; PS00666; DHDPS_2; 1.
KW Lyase; Diaminopimelate biosynthesis; Lysine biosynthesis;
KW Complete proteome.
FT ACT_SITE 166 166 BY SIMILARITY.
SQ SEQUENCE 300 AA; 32345 MW; 42EFA59FE0D8883 CRC64;
Query Match 20.2%; Score 234.5; DB 1; Length 300;
Best Local Similarity 32.2%; Pred. No. 1.2e-11;
Matches 68; Conservative 42; Mismatches 94; Indels 7; Gaps 6;
Qy 7 KLOGLVAATTPMTENGESVITGVYDVKVKEQGVKNIFVNGTTGEGLSLSVSRQW 65
Db 4 QIEGVIPALITPFTDGLKGINEEGLRENVSRL-LAAGVHVVPVAGTTGESSTLSHAHRR 62
Qy 66 VAESWVTGKDKLDQVIVHVGALSKEQELAQHAAEIGADGIAVIAPFLKPTWKDILI 125
Db 63 VIETWDEVNGKV-PVLAGAGSNSTREALELSTVAEDVGADAILSWPYNKPQEGJFI 121
Qy 126 NFLKEVAAAAPALPFYYHIPALTGVKIRAEELLDGILDKIPTFQGLKFSDDTLDFGQCV 185
Db 122 HFSK--IAEVECEIILYNPSTRGCALEPTAAK-LAEYSYHVGKESGD-LDVVQR 177
Qy 186 VDQNRQQQFAFLFGVDEQLLSALVMGATGAV 216

```

```

Db 178 FIETPPDFILLGSGVDELTL2PLANGGVGI 208
RESULT 14
DAPA_METH
ID DAPA_METH STANDARD; PRT; 283 AA.
AC O26892;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydrodipicolinate synthase (EC 4.2.1.52) (DHDP5).
GN DAPA OR MTH801.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
CX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Slakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.-J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- CATALYTIC ACTIVITY: L-aspartate 4-semialdehyde + pyruvate =
CC dihydrodipicolinate + 2 H(2)O.
CC -1- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
CC semialdehyde; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the DHDP5 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A69207; A69207.
CC PIR; A69207; A69207.
CC HSSP; PC5640; 1DHP.
CC HAMAP; MF_00418; -; 1.
CC InterPro; IPR005263; DAPA.
CC Pfam; PF00701; DHDP5; 1.
CC PRINTS; PR00146; DHPICSNTHASE.
CC ProDom; PD001859; DHDP5; 1.
CC TIGRFAMs; TIGR00674; dapa; 1.
CC PROSITE; PS00665; DHDP5_1; 1.
CC PROSITE; PS00666; DHDP5_2; 1.
CC Lyase; Diaminopimelate biosynthesis; Lysine biosynthesis;
KW Complete proteome.
FT ACT SITE 162 162 BY SIMILARITY.
SQ SEQUENCE 283 AA; 30362 MW; 8F71C2E574D83B64 CRC64;

Query Match 20.0%; Score 231.5; DB 1; Length 283;
Best Local Similarity 27.3%; Pred. No. 2e-11;
Matches 62; Conservative 52; Mismatches 90; Indels 23; Gaps 7;

QY 7 KLQGLVAAITPTENGSEINFSVIGQVDYLVKEQGVQNFVNGTTGEGLSVSRQV 66
Db 2 KEGTVAWVTPEDDVVDVDEAGLRNINYL-ENGVDGLVAGTIGESATITHEQRRM 60
QY 67 AEWVTGKDKLDQWIIHVGALSLSKESQBLAQAHAIGADGATVAPFLKPKWTDILIN 126
Db 61 IDILVDEVNGRV-KTVAGAGSNSREAMGLVEYADAGADAALVITPYNKPQPHGLIEH 119

```

```

QY 127 R--LKEVAAAAPALPFYVYHIPALTGVKIRAEEL-----LDGIL---DKIPTFOGLKPSD 176
Db 120 YTMLEEEA-----DPLIIYNVPSRTGTIDVDTVAELAKLGLIGIKASPDLDKVSMLR 175
QY 177 TDLIDFGQCVDPNRQQQFAFLFGVDEQLLSALVMGATGAVGSFVSRD 223
Db 176 SRLMDLG-----LDDFTVLSGNDLTLPLMISMGAGVISVWAVD 215

RESULT 15
DAPA_METMA
ID DAPA_METMA STANDARD; PRT; 291 AA.
AC Q8PX17;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydrodipicolinate synthase (EC 4.2.1.52) (DHDP5).
GN DAPA OR MM1201.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
CX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goei / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Waez A., Baumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boencke M., Steckel S.,
RA Bhattacharya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsäus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
RT transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC -1- CATALYTIC ACTIVITY: L-aspartate 4-semialdehyde + pyruvate =
CC dihydrodipicolinate + 2 H(2)O.
CC -1- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
CC semialdehyde; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the DHDP5 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A6913350; AAM30897.1; -.
CC HAMAP; MF_00418; -; 1.
CC InterPro; IPR005263; DAPA.
CC InterPro; IPR002220; DHDP5.
CC Pfam; PF00701; DHDP5; 1.
CC PRINTS; PR00146; DHPICSNTHASE.
CC ProDom; PD001859; DHDP5; 1.
CC TIGRFAMs; TIGR00674; dapa; 1.
CC PROSITE; PS00665; DHDP5_1; 1.
CC PROSITE; PS00666; DHDP5_2; 1.
CC Lyase; Diaminopimelate biosynthesis; Lysine biosynthesis;
KW Complete proteome.
FT ACT SITE 159 159 BY SIMILARITY.
SQ SEQUENCE 291 AA; 30704 MW; 45BF0B0D017FFA26 CRC64;

Query Match 19.5%; Score 226; DB 1; Length 291;
Best Local Similarity 28.8%; Pred. No. 5.7e-11;
Matches 60; Conservative 44; Mismatches 96; Indels 8; Gaps 4;

QY 9 QGLVAAITPTENGSEINFSVIGQVDYLVKEQGVQNFVNGTTGEGLSVSRQV 68
Db 3 EGNPALITPFDKDRIDREGLRNIAP-VEEGVSGIVPCGTTGEGSATLSAAHEVID 61

```

Qy	69	EWVTGKXKLDQVITHGALSLSKESQELAQHAEIGADSI	AVIAPFLKPTKDILINFL	128
Db	62	IAVECSK--VPVIAGTGSNNTGEALQFTKHAADAGVDG	VLLISPYNKNPNPAGLLAHFK	118
Qy	129	KEVAAAAPALPFYVYHIPALTGVKIRAEELLOGILD	KIPTFOGLKFSDTDLIDFGQCVDQ	188
Db	119	X--IAEAVDIEWILYNVFSRTGQDMPVDVIVE--L	AKVENIVGKEASGNAKVSQILEN	174
Qy	189	NRQOQFAPLFGVDEQLLSALVVGATGAV		216
Db	175	TMDDDFVVLSCEDGTLPLISMGGRGVI		202

Search completed: May 6, 2004, 09:04:58
Job time : 7.04821 secs

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:59:19 ; Search time 22.5128 Seconds
(without alignments)
3223.466 Million cell updates/sec

Title: US-09-930-440b-2
Perfect score: 1159
Sequence: 1 MAPPKKQLGLVAATITPMT.....GATGAVGFSVRSJLSTLSN 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Result No.	Score	Query Match	Length	DB ID	Description
1	1104.5	95.3	320	4 Q9BXD5	Q9bxds5 homo sapien
2	1032.5	89.1	319	6 Q9BEC7	Q9bec7 sus scrofa
3	998.5	86.2	320	11 Q9DC99	Q9dcj9 mus musculu
4	443	38.2	297	16 Q8ZCG7	Q8zcg7 yersinia pe
5	441	38.1	299	16 Q8D0X1	Q8d0x1 yersinia ps
6	430	37.1	298	16 Q9KR67	Q9kr67 vibrio chol
7	421.5	36.4	322	16 Q7UUE0	Q7ueu0 rhodospirell
8	399	34.4	299	16 Q8D617	Q8d617 vibrio vuln
9	273	23.6	306	2 Q8RR42	Q8rr42 streptococc
10	256	22.1	305	16 Q97P53	Q97pf53 streptococc
11	252.5	21.8	305	16 Q97Q96	Q97q96 streptococc
12	251.5	21.7	294	17 Q9UZ94	Q9uz94 pyrococcus
13	248.5	21.4	318	5 Q27818	Q27818 trichomonas
14	247	21.3	305	16 Q8E7V7	Q8e7v7 streptococc
15	247	21.3	305	16 Q8E2E9	Q8e2e9 streptococc
16	243	21.0	226	2 Q32375	Q32375 clostridium

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ID	Q9BXD5	PRELIMINARY;	PRT;	320 AA.
AC	Q9BXD5;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	Clorf13			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=2121827; PubMed=11318611;			
RA	Sood R., Bonner T.I., Malakowska I., Stephan D.A., Robbins C.M.,			
RA	Connors T.D., Morgenbesser S.D., Su K., Faruque M.U., Pinkett H.,			
RA	Graham C., Baxevasis A.D., Klinger K.W., Landes G.M., Trent J.M.,			
RA	Carpten J.D.;			
RT	*Cloning and characterization of 13 novel transcripts and the human			
RT	RG88 gene from the lq25 region encompassing the hereditary prostate			
RT	cancer (hpc1) locus";			
RL	Genomics 73:211-222(2001).			
DR	EMBL; AF338436; AAK25795.1; -			
DR	Genew; HGNC:16781; Clorf13.			
DR	InterPro; IPR002220; DHDP5.			
DR	Pfam; PF00701; DHDP5; 1.			
DR	PRINTS; PR00146; DHDP5NTBASE.			
DR	ProDom; PD001859; DHDP5; 1.			
SQ	SEQUENCE 320 AA; 35162 MW; FCOEBA9B05FE9E62 CRC64;			

Query Match	95.3%;	Score 1104.5;	DB 4;	Length 320;
Best Local Similarity	91.0%;	Pred. No. 4.5e-86;		
Matches 222;	Conservative	2;	Mismatches 5;	Indels 15;
Gaps	1;			

QY	1	MAFPKKKQLGLVAATITPMTENGEINFSVIGQVVDLVKQGVKNIFVNGTTGEGLSLV	60
DB	1	MAFPKKKQLGLVAATITPMTENGEINFSVIGQVVDLVKQGVKNIFVNGTTGEGLSLV	60

ALIGNMENTS

RESULT 1	Q9BXD5	PRELIMINARY;	PRT;	320 AA.
AC	Q9BXD5;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	Clorf13			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=2121827; PubMed=11318611;			
RA	Sood R., Bonner T.I., Malakowska I., Stephan D.A., Robbins C.M.,			
RA	Connors T.D., Morgenbesser S.D., Su K., Faruque M.U., Pinkett H.,			
RA	Graham C., Baxevasis A.D., Klinger K.W., Landes G.M., Trent J.M.,			
RA	Carpten J.D.;			
RT	*Cloning and characterization of 13 novel transcripts and the human			
RT	RG88 gene from the lq25 region encompassing the hereditary prostate			
RT	cancer (hpc1) locus";			
RL	Genomics 73:211-222(2001).			
DR	EMBL; AF338436; AAK25795.1; -			
DR	Genew; HGNC:16781; Clorf13.			
DR	InterPro; IPR002220; DHDP5.			
DR	Pfam; PF00701; DHDP5; 1.			
DR	PRINTS; PR00146; DHDP5NTBASE.			
DR	ProDom; PD001859; DHDP5; 1.			
SQ	SEQUENCE 320 AA; 35162 MW; FCOEBA9B05FE9E62 CRC64;			

Q83ca6	coxiella bu
Q97wf0	sulfolobus
Q97wf2	sulfolobus
Q58577	pyrococcus
Q81cd9	bacillus ce
Q8u319	pyrococcus
Q81ph3	bacillus an
Q81w7	bacillus ce
Q819z9	bacillus ce
Q8k9p0	streptococc
Q9a1i5	streptococc
Q8p2p3	streptococc
Q8h7z5	phytophthor
Q8g28	bradyrhizob
Q96y33	sulfolobus
Q7ua33	synechococc
Q7umj1	rhodospirell
Q7v930	prochloroco
Q8nht2	pseudomonas
Q8emh5	oceanobacil
Q97u22	sulfolobus
Q9jrj8	methyllobaci
Q9hj17	thermoplasma
Q8fjw9	escherichia
Q7vm87	haemophilus
Q82sd7	nitrosomona
Q7vx28	bordetella
Q7w884	bordetella
Q7x387	escherichia

```

QY 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGHDGLAVIAPFFLKPT 120
DB 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGHDGLAVIAPFFLKPT 120
QY 121 KDILINFLKEVAAPALPFFYYHHPALTGKIRAEELLDGILDKIPTFQGLKESDIDL 180
DB 121 KDILINFLKEVAAPALPFFYYHHPALTGKIRAEELLDGILDKIPTFQGLKESDIDL 180
QY 181 DFGQCVQNRQQQFAFLFGVDEQQLSALVNGATGAVGS-----FVSRDLS 225
DB 181 DFGQCVQNRQQQFAFLFGVDEQQLSALVNGATGAVGS-----FVSRDLS 225
QY 226 TLIS 229
DB 241 SALNH 244

RESULT 2
Q9BEC7 PRELIMINARY; PRT; 319 AA.
AC Q9BEC7; (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 24, Last annotation update)
DE Acylneuraminase lyase.
EN NPL
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Bruse P., Travling C., Schauer R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ271330; CAC27797.1; -.
DR GO; GO:0016829; F:lyase activity; IEA.
DR InterPro; IPR002220; DHDP5.
DR Pfam; PF00731; DHDP5; 1.
DR PRINTS; PR00146; DHDP5; 1.
DR ProDom; PD001859; DHDP5; 1.
KW Lyase.
SQ SEQUENCE 319 AA; 35062 MW; C8989459F68674A7 CRC64;

Query Match
Best Local Similarity 84.1%; Score 1032.5; DB 6; Length 319;
Matches 206; Conservative 13; Mismatches 11; Indels 15; Gaps 1;

QY 1 MAFPKKKLQGLVAATITPMTENGINFVIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
DB 1 MAFPKKKLQGLVAATITPMTENGINFVIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
QY 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGHDGLAVIAPFFLKPT 120
DB 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGHDGLAVIAPFFLKPT 120
QY 121 KDILINFLKEVAAPALPFFYYHHPALTGKIRAEELLDGILDKIPTFQGLKESDIDL 180
DB 121 KDILINFLKEVAAPALPFFYYHHPALTGKIRAEELLDGILDKIPTFQGLKESDIDL 180
QY 181 DFGQCVQNRQQQFAFLFGVDEQQLSALVNGATGAVGS-----FVSRDLS 225
DB 181 DFGQCVQNRQQQFAFLFGVDEQQLSALVNGATGAVGS-----FVSRDLS 225
QY 226 TLISN 230
DB 241 SALNH 245

RESULT 3
Q9BEC7 PRELIMINARY; PRT; 320 AA.
AC Q9BEC7; (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 24, Last annotation update)
DE Acylneuraminase lyase.
EN NPL
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Bruse P., Travling C., Schauer R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ271330; CAC27797.1; -.
DR GO; GO:0016829; F:lyase activity; IEA.
DR InterPro; IPR002220; DHDP5.
DR Pfam; PF00731; DHDP5; 1.
DR PRINTS; PR00146; DHDP5; 1.
DR ProDom; PD001859; DHDP5; 1.
KW Lyase.
SQ SEQUENCE 319 AA; 35062 MW; C8989459F68674A7 CRC64;

Query Match
Best Local Similarity 84.1%; Score 1032.5; DB 6; Length 319;
Matches 206; Conservative 13; Mismatches 11; Indels 15; Gaps 1;

QY 1 MAFPKKKLQGLVAATITPMTENGINFVIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
DB 1 MAFPKKKLQGLVAATITPMTENGINFVIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
QY 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGHDGLAVIAPFFLKPT 120
DB 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGHDGLAVIAPFFLKPT 120
QY 121 KDILINFLKEVAAPALPFFYYHHPALTGKIRAEELLDGILDKIPTFQGLKESDIDL 180
DB 121 KDILINFLKEVAAPALPFFYYHHPALTGKIRAEELLDGILDKIPTFQGLKESDIDL 180
QY 181 DFGQCVQNRQQQFAFLFGVDEQQLSALVNGATGAVGS-----FVSRDLS 225
DB 181 DFGQCVQNRQQQFAFLFGVDEQQLSALVNGATGAVGS-----FVSRDLS 225
QY 226 TLISN 230
DB 241 SALNH 245

```

```

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 36-0033B02Rik protein (RIKEN cDNA 0610033B02 gene)
DE (N-acetylneuraminase pyruvate lyase).
EN NPL OR 0610033B02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino Y., Konno H., Adachi J., Fukuda S.,
RA Arakawa T., Hara A., Fukunishi Y., Kondo S., Yamamoto I.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kasakawa T., Saito R.,
RA Saito T., Okazaki Y., Gojobori T., Soto H., Kasakawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki K., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli K., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Monbaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD; TISSUE=Thymus;
MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RL Nature 420:563-573(2002).
RA "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
DR EMBL; AK002734; BAB22314.1; -.
DR EMBL; BC022734; AAB22734.1; -.
DR EMBL; AK088859; BAC40618.1; -.
DR MGD; MGI:1921341; Npl.
DR InterPro; IPR002220; DHDP5.
DR Pfam; PF00701; DHDP5; 1.
DR PRINTS; PR00146; DHDP5; 1.
DR ProDom; PD001859; DHDP5; 1.
SQ SEQUENCE 320 AA; 35130 MW; 9426AD7CC8438468 CRC64;

Query Match
Best Local Similarity 86.2%; Score 998.5; DB 11; Length 320;
Matches 198; Conservative 17; Mismatches 14; Indels 15; Gaps 1;

QY 1 MAFPKKKLQGLVAATITPMTENGINFVIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
DB 1 MAFPKKKLQGLVAATITPMTENGINFVIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
QY 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGHDGLAVIAPFFLKPT 120
DB 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGHDGLAVIAPFFLKPT 120
QY 121 KDILINFLKEVAAPALPFFYYHHPALTGKIRAEELLDGILDKIPTFQGLKESDIDL 180
DB 121 KDILINFLKEVAAPALPFFYYHHPALTGKIRAEELLDGILDKIPTFQGLKESDIDL 180
QY 181 DFGQCVQNRQQQFAFLFGVDEQQLSALVNGATGAVGS-----FVSRDLS 225
DB 181 DFGQCVQNRQQQFAFLFGVDEQQLSALVNGATGAVGS-----FVSRDLS 225
QY 226 TLISN 230
DB 241 SALNH 245

```


Thu May 13 11:53:01 2004

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative acetylneuraminidase lyase.
 GN Y1458.
 OS Versinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OC NCBI_TaxID=632;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Mediaevalis;
 EX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.B., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of Yersinia pestis KIM.";
 RL J. Bacteriol. 184:4501-4611(2002).
 DR EMBL; AB013749; AM85029.1; -;
 DR GO; GO:0016829; F-lyase activity; IEA.
 DR InterPro; IPR002220; DHDPS.
 DR Pfam; PF00701; DHDPS; 1.
 DR PRINTS; PR00146; DHDPS; 1.
 DR PRODOM; PD001859; DHDPS; 1.
 DR PROSITE; PS00665; DHDPS_1; 1.
 DR LYASE.
 SQ SEQUENCE 299 AA; 32179 MW; 846EBF6165791575 CRC64;
 KW Lyase.
 QY 6 KKLQGLVAATITPTWENGELNFSVIGQYDYLKVEQGVNIFVNGTGTGEGLSLSVSRQ 65
 DB 4 KKLTLGLIAAPHTPFDEQGEVNPVIDQIAEHLND-GVKGYYVCGTGTGEGHCSVDERKK 62
 QY 66 VAEWTKGDKLDQVLIHVGALSLKESQELAAHAEIGADGLAVIAPFFLKPTKDI 125
 DB 63 IAEWYNAAGKL-SITLHTGALSINDAVLSRHAETLDIFATSAIGCFKFGNLDLI 121
 QY 126 NFLKEVAAAAPALPFYHYHHPALTVKIRAEELLDGLDKIPTFGSKFSDTLLDFGQC 185
 DB 122 AYCQAIATAAPSKGFYYH-SGMSGVNLDMQELIKAESKIPNLGKFNADLYEFQRC 180
 QY 186 VQNRQOQAFALFGVDEQLLSALVMGATGAVS 218
 DB 181 L-RVSGGKFDIPFVDEHLPGGLAVGAVGS 212
 RESULT 5
 Q9KR67 PRELIMINARY; PRT; 298 AA.
 ID Q9KR67
 AC Q9KR67
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE N-acetylneuraminidase lyase, putative.
 GN VCI776.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OC NCBI_TaxID=666;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=E1 Tor N16961 / Serotype O1;
 EX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.J.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

QY 181 DFGQCVQDNRQOQAFALFGVDEQLLSALVMGATGAVS-----FVSRDL 225
 DB 181 DFGQCVQDNRQOQAFALFGVDEQLLSALVMGATGAVS-TYNLGGKTNOMLEAFQKDLA 240
 QY 226 TLLS 229
 DB 241 SALS 244
 RESULT 4
 Q8ZG37 PRELIMINARY; PRT; 297 AA.
 ID Q8ZG37
 AC Q8ZG37
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Probable N-acetylneuraminidase lyase (EC 4.1.3.3).
 GN YP03024.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OC NCBI_TaxID=632;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 EX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 DR EMBL; AJ14155; CAC9266.1; -;
 DR PIR; AG0367; AG0367.
 DR GO; GO:0016829; F-lyase activity; IEA.
 DR GO; GO:0008747; F-N-acetylneuraminidase lyase activity; IEA.
 DR InterPro; IPR002220; DHDPS.
 DR Pfam; PF00701; DHDPS; 1.
 DR PRINTS; PR00146; DHDPS; 1.
 DR PRODOM; PD001859; DHDPS; 1.
 DR PROSITE; PS00665; DHDPS_1; 1.
 DR LYASE; Complete proteome.
 SQ SEQUENCE 297 AA; 31950 MW; 57C1D709D25514C5 CRC64;
 KW Lyase.
 QY 6 KKLQGLVAATITPTWENGELNFSVIGQYDYLKVEQGVNIFVNGTGTGEGLSLSVSRQ 65
 DB 2 KKLTLGLIAAPHTPFDEQGEVNPVIDQIAEHLND-GVKGYYVCGTGTGEGHCSVDERKK 60
 QY 66 VAEWTKGDKLDQVLIHVGALSLKESQELAAHAEIGADGLAVIAPFFLKPTKDI 125
 DB 61 IAEWYNAAGKL-SITLHTGALSINDAVLSRHAETLDIFATSAIGCFKFGNLDLI 119
 QY 126 NFLKEVAAAAPALPFYHYHHPALTVKIRAEELLDGLDKIPTFGSKFSDTLLDFGQC 185
 DB 120 AYCQAIATAAPSKGFYYH-SGMSGVNLDMQELIKAESKIPNLGKFNADLYEFQRC 178
 QY 186 VQNRQOQAFALFGVDEQLLSALVMGATGAVS 218
 DB 179 L-RVSGGKFDIPFVDEHLPGGLAVGAVGS 210
 RESULT 5
 Q8D0X1 PRELIMINARY; PRT; 299 AA.
 ID Q8D0X1
 AC Q8D0X1

```
RA Salzberg S.B., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RJ Nature 406:477-483 (2000).
DR EMBL; AB004255; AAF94925.1; -.
DR FIR; B82157; B82157.
DR TIGR; VC1776; -.
DR GO; GO:0016829; F-lyase activity; IEA.
DR InterPro; IPR002220; DHDP5.
DR PRINTS; PF00701; DHDP5; 1.
DR ProDom; PD001859; DHDP5; 1.
KW Lyase; Complete proteome.
SQ SEQUENCE 298 AA; 32221 MW; 4359567DC749BB6D CRC64;

Query Match 37.1%; Score 430; DB 16; Length 298;
Best Local Similarity 44.1%; Pred. No. 1.7e-28;
Matches 94; Conservative 39; Mismatches 76; Indels 4; Gaps 4;

QY 6 KKQGLVAAATTPMTENGESVIGQYDVYLVKEQGVKNIFVNGTTGEGSLSVSERRQ 65
Db 2 KKLGLIAAPHTFFKKNKFNRAIDQIAELL-EGVGKGYVCGTTGEGIHGCSVERKA 60

QY 66 VAEWWTGKDKLDQVTHVHGSLSKESQELAHAAEIGADGIAVIAPFFKPKWTKDILI 125
Db 61 IASRWKAVDGLD-VILHTGALSIVDTLHTHAETLDIFATSAIGPCFFKPGSVDDL 119

QY 126 NELKEVAAAAPALPFYYHHPALTGKIRAEELDGLDKIPTFGQLKFSDDLDFGQC 185
Db 120 EYCAQVAAAAPSKGFYYH-SGSGVNLIDLEOFILKGEQRIPLNYGAKFNADLYEQRC 178

QY 186 VDQNRQQQAFLEFGVDQQLSALVMGATGAVGS 218
Db 179 V-RVSNRKFDELPFGVDEFLPAGLAVGAGVGS 210

RESULT 7
Q7JUEO PRELIMINARY; PRT; 322 AA.
AC Q7JUEO;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25; Last sequence update)
DE Probable N-acetylneuraminatase lyase (EC 4.1.3.3).
GN NANA OR RH3352.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=20735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RC "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294138; CAD73140.1; -.
KW Lyase; Complete proteome.
SQ SEQUENCE 322 AA; 34857 MW; 02BAC02AB21AA68C CRC64;

Query Match 36.4%; Score 421.5; DB 16; Length 322;
Best Local Similarity 43.7%; Pred. No. 9.8e-28;
Matches 93; Conservative 41; Mismatches 76; Indels 3; Gaps 3;

QY 6 KKQGLVAAATTPMTENGESVIGQYDVYLVKEQGVKNIFVNGTTGEGSLSVSERRQ 65
Db 15 RKLGLIAATPMPKACGDLNDVVPAMVEKLLD-GVSGVCGTSGEGLSRTREQL 73

QY 66 VAEWWTGKDKLDQVTHVHGSLSKESQELAHAAEIGADGIAVIAPFFKPKWTKDILI 125
```

```
Db 74 VASAFVNAAGRV-PVIVQGENSLAEAOELAKHAQOIGASALSATCSYFKVASQSLT 132
QY 126 NFKVEVAAAAPALPFYYHHPALTGKIRAEELDGLDKIPTFGQLKFSDDLDFGQC 185
Db 133 LCMQELAAAAAPETFFYYHHPVLTGSSIDMVFEFTTHADEAIPITVLGKTYDTMLFEFQRC 192
QY 186 VDQNRQQQAFLEFGVDQQLSALVMGATGAVGS 218
Db 193 LELS-NRKEDVWVWGCDEMMLGATATGARAAGS 224

RESULT 8
Q8D617 PRELIMINARY; PRT; 239 AA.
AC Q8D617;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23; Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24; Last annotation update)
DE Dihydrodipicolinate synthase/N-acetylneuraminatase lyase.
GN VV20730.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (JBC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEC16810; AA007662.1; -.
DR GO; GO:0016829; F-lyase activity; IEA.
DR InterPro; IPR002220; DHDP5.
DR Pfam; PF00701; DHDP5; 1.
DR PRINTS; PR00146; DHPICSNTHASE.
DR ProDom; PD001859; DHDP5; 1.
KW Lyase; Complete proteome.
SQ SEQUENCE 239 AA; 32199 MW; 69CEC84008AB211A CRC64;

Query Match 34.4%; Score 399; DB 16; Length 299;
Best Local Similarity 41.5%; Pred. No. 7.4e-26;
Matches 88; Conservative 40; Mismatches 80; Indels 4; Gaps 4;

QY 7 KQGLVAAATTPMTENGESVIGQYDVYLVKEQGVKNIFVNGTTGEGSLSVSERRQV 66
Db 4 KKLGLIAAPHTFFDNNQVRYVIDQIAALLI-EQGVGAYVCGTTGEGIHGCSVERKAI 62

QY 67 AEEWWTGKDKLDQVTHVHGSLSKESQELAHAAEIGADGIAVIAPFFKPKWTKDILI 126
Db 63 AERWVKAADGKLD-IILHTGALSIVDTLHTHAETLDILATSAIGPCFFKPSVADLVN 121

QY 127 FLKEVAAAAPALPFYYHHPALTGKIRAEELDGLDKIPTFGQLKFSDDLDFGQC 186
Db 122 YCAQTAAAPSKGFYYH-SGSGVNLIDLEOFILKGEQRIPLNSGAKFNVDLYEQRAL 180

QY 187 DQNRQQQAFLEFGVDQQLSALVMGATGAVGS 218
Db 181 -RVSNRKFDELPFGVDEFLPAGLAVGAGVGS 211

RESULT 9
Q8RR42 PRELIMINARY; PRT; 305 AA.
AC Q8RR42;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21; Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25; Last annotation update)
DE Putative N-acetylneuraminatase lyase.
GN NANA.
OS Streptococcus suis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
```

Thu May 13 11:53:01 2004

```

OC Streptococcus.
OX NCBI_TaxID=1307;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=10581; PubMed=11889115;
RX MEDLINE=21386400; Sekizaki T.;
RA Takamatsu D., Osaki M., Sekizaki T.;
RT "Evidence for Lateral Transfer of the sulfolysin Gene Region of
RL Streptococcus suis.";
RU J. Bacteriol. 184:2050-2057(2002).
DR EMEL; AB071365; BAB85970.1; -
DR GO: GO:0016829; F:lyase activity; IEA.
DR InterPro; IPR002220; DHDPS.
DR InterPro; IPR000437; Prok_lipoprot.S.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHDPS; 1.
DR PRODOM; PD001859; DHDPS; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Lyase.
SQ SEQUENCE 306 AA; 33402 MW; 038A876FE750356D CRC64;

Query Match 23.6%; Score 273; DB 2; Length 306;
Best Local Similarity 33.5%; Pred. No. 4.4e-15;
Matches 72; Conservative 41; Mismatches 89; Indels 12; Gaps 6;

QY 7 KXQGLVAATITPMTENGINFVSIGQYDYLKQGVKNIFVNGTGBGLSLSVSERRQV 66
DB 6 KXHGVIPIAFYACVDRAGEISSEKVKLVQYFT-DKGVQGLYVNGSGSGECIQSVADRKQI 64

QY 67 AEENWTKGDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFFLPKPTXDLIN 126
DB 65 LEAVMEVAKGKL-TIIHVACNNKDSVLAHSEELGVDATIAIPPIYFR-LPEYSVA 122

QY 127 FLKEVAAAPALPFYVYHIPALTGVKIRAE---ELLGDILDKIPTFGQLKFSDTLLDFG 183
DB 123 YWNGISAAAPNTDFIYNIPQLAGVSLTSLYKEML-----KNPRVGVKNSSMPVQDI 177

QY 184 QCVQDQNRQQQFAFLFGVDEQLLSALVNGATGAVG 217
DB 178 TFVTLGGEDYVWF-NGPDEQFLGRLMGARAGIG 210

RESULT 10
Q972F3 PRELIMINARY; PRET; 305 AA.
AC Q97PF3;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-acetylneuraminase lyase, putative.
GN SP1676.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapf E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMEL; AE007431; AAK75427.1; -.
DR PIR; B95154; B95154.
DR TIGR; SP1329; -.
DR GO: GO:0016829; F:lyase activity; IEA.
DR InterPro; IPR002220; DHDPS.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHDPS; 1.
DR PRODOM; PD001859; DHDPS; 1.
KW Lyase. Complete proteome.
SQ SEQUENCE 305 AA; 33562 MW; 5F8A661DD5C86ED7 CRC64;

Query Match 21.8%; Score 252.5; DB 16; Length 305;
Best Local Similarity 31.1%; Pred. No. 2.5e-13;
Matches 73; Conservative 37; Mismatches 104; Indels 21; Gaps 6;

QY 7 KXQGLVAATITPMTENGINFVSIGQYDYLKQGVKNIFVNGTGBGLSLSVSERRQV 66

```

```

DR GO: GO:0016829; F:lyase activity; IEA.
DR InterPro; IPR002220; DHDPS.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHDPS; 1.
DR PRODOM; PD001859; DHDPS; 1.
KW Lyase; Complete proteome.
SQ SEQUENCE 305 AA; 33425 MW; 2912FAAD6DC720A7 CRC64;

Query Match 22.1%; Score 256; DB 16; Length 305;
Best Local Similarity 31.2%; Pred. No. 1.3e-13;
Matches 67; Conservative 42; Mismatches 94; Indels 12; Gaps 6;

QY 6 KXQGLVAATITPMTENGINFVSIGQYDYLKQGVKNIFVNGTGBGLSLSVSERRQV 65
DB 5 KXHGVIPIAFYACVDDQGEVSPERTRALQYFT-DKGVQGLYVNGSGSGECIQSVDRKL 63

QY 66 VAEENWTKGDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFFLPKPTXDLIN 125
DB 64 ILEENMAVAKGKL-TIIHVACNNKDSVLAHSEELGVDATIAIPPIYFR-LPEYSVA 121

QY 126 NFLKEVAAAPALPFYVYHIPALTGVKIRAE---ELLGDILDKIPTFGQLKFSDTLLDF 182
DB 122 KYWNDISSAAPNTDFIYNIPQLAGVSLTSLYKEML-----KNPRVGVKNSSMPVQDI 176

QY 183 QCVQDQNRQQQFAFLFGVDEQLLSALVNGATGAVG 217
DB 177 QTFVSLGGEDYVWF-NGPDEQFLGRLMGARAGIG 210

RESULT 11
Q97Q96 PRELIMINARY; PRET; 305 AA.
AC Q97Q96;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-acetylneuraminase lyase.
GN SP1329.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapf E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMEL; AE007431; AAK75427.1; -.
DR PIR; B95154; B95154.
DR TIGR; SP1329; -.
DR GO: GO:0016829; F:lyase activity; IEA.
DR InterPro; IPR002220; DHDPS.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHDPS; 1.
DR PRODOM; PD001859; DHDPS; 1.
KW Lyase. Complete proteome.
SQ SEQUENCE 305 AA; 33562 MW; 5F8A661DD5C86ED7 CRC64;

Query Match 21.8%; Score 252.5; DB 16; Length 305;
Best Local Similarity 31.1%; Pred. No. 2.5e-13;
Matches 73; Conservative 37; Mismatches 104; Indels 21; Gaps 6;

QY 7 KXQGLVAATITPMTENGINFVSIGQYDYLKQGVKNIFVNGTGBGLSLSVSERRQV 66

```

Thu May 13 11:53:01 2004

us-09-930-440b-2.rspt

```

Db      6 KYGVIPAFACYDENGESQDRKSLGVYFI-DKGVKGIYVNGSGEGCIYVSVDKRI 64
QY      67 AEWTKGKQKLDQVLIHVGALSKEQELAGADGAGIAPFFLKPTWKDILIN 126
Eb      65 IFAWVEVAKGKI-TVINELACNNTKDSIELAKHSESVGVDAJAAIPPIYFK-LPEYSIAA 122
QY      127 FLKEVAAAAPALPFYYHYHIPALTGVKIRAEELLDGILKIPFQGLKFSDDTLDDPGQCV 186
Db      123 YNNAMSEAAASNTDFIYINIPLAGVALTGS--LYATMRQNPRIYGVKNSMPVQDIDQXVF 180
QY      187 DQNRQOQAFELGVDEQLLSALVWGATGAVS-----FVSRLST 226
Db      181 AAGGEDYVVF-NGEDCYLGGRLMGAEAGIGYGVNPDILFLKLESLIQERLDLT 234

RESULT 12
Q9UZ94 PRELIMINARY; PRT; 294 AA.
AC Q9UZ94;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Dihydrodipicolinate synthase (DAPA).
GN DAPA OR PYRAB12600 OR PAB0832.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=GBS / Orsay;
RC Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EXBL; AJ248297; CAB50165.1; -.
DR PIR; H75033; H75033.
DR InterPro; IPR002220; DHPS.
DR Pfam; PF00701; DHPS; 1.
DR PRINTS; PR00146; DHPICNTHASE.
DR ProDom; PD001859; DHPS; 1.
KW Complete Proteome.
SQ SEQUENCE 294 AA; 32913 MW; 7FA3D40BF9E9FAC2 CRC64;

Query Match 21.7%; Score 251.5; DB 17; Length 294;
Best Local Similarity 30.1%; Pred. No. 2.9e-13;
Matches 65; Conservative 49; Mismatches 81; Indels 21; Gaps 7;

QY 8 LQGLVAATITPMTENGEINFSVIGQVDLYKEQGVKNIFVNTTGEGLSLSVSRROVA 67
Db 1 MEGVIVELVTPFREDYSIDFALWHEHREL--EEKGVHGVFNSTTGETSINTDERKLA 59
QY 68 EEW-VTKGDKLDQVLIHVGALSKEQELAGADGAGIAPFFLKPTWKDILIN 125
Db 60 EKGREITSG-----MYLVGTSTSTLEVLISRAEDIGADGIV-VAPYCKLKDEILK 114
QY 126 NPLKEVAAAAPALPFYYHYHIPALTGVKIRAEELLDGILKIPFQGLKFSDDTLDD 181
Db 115 HF--SMVAERVDIPITVVAIPSC-----ANPVPVDIIRKVSLEYSNIIGKVASVDSJCY 166
QY 182 FGQCYD-QNRQOQAFELGVDEQLLSALVWGATGAV 216
Db 167 -QELIEVKEERKDFRVFTGLDQFYLSLLGGDGGI 202

RESULT 13
Q27818 PRELIMINARY; PRT; 318 AA.
AC Q27818;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

```

```

DE N-acetylneuraminase lyase (EC 4.1.1.3.3).
GN NANA.
OS Trichomonas vaginalis.
OC Eukaryota; Parabasalidea; Trichomonadida; Trichomonadidae;
OC Trichomonadinae; Trichomonas.
OX NCBI_TaxID=5722;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=clinical isolate #202;
RX MEDLINE=97077447; PubMed=8920014;
RA Meysick K.C., Dimock K., Garber G.B.;
RT "Molecular characterization and expression of a N-acetylneuraminase
lyase gene from Trichomonas vaginalis.";
RL Mol. Biochem. Parasitol. 76:289-292(1996).
DR EMBL; U35878; AAB42182.1; -.
DR HSSP; P06995; LFDY.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0018299; F:lyase activity; IEA.
DR GO; GO:0008747; F:N-acetylneuraminase lyase activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR002220; DHPS.
DR InterPro; IPR005264; Nana.
DR InterPro; IPR005829; Sug_transporter.
DR Pfam; PF00701; DHPS; 1.
DR PRINTS; PR00146; DHPICNTHASE.
DR ProDom; PD001859; DHPS; 1.
DR TIGRfams; TIGR00683; nana; 1.
DR PROSITE; PS00665; DHPS_1; 1.
DR PROSITE; PS00666; DHPS_2; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Lyase.
SQ SEQUENCE 318 AA; 35083 MW; 25F0244C298D593D CRC64;

Query Match 21.4%; Score 248.5; DB 5; Length 318;
Best Local Similarity 30.0%; Pred. No. 5.8e-13;
Matches 64; Conservative 47; Mismatches 95; Indels 7; Gaps 5;

QY 5 KKLQGLVAATITPMTENGEINFSVIGQVDLYKEQGVKNIFVNTTGEGLSLSVSRRO 65
Db 27 KSLKGLFSALLVSPNEDGTINEKGLREIVRYNIDKMKIDGLYGVGSGTENFELTEKKQ 86
QY 66 VAEWVTGDKLDQVLIHVGALSKEQELAGADGAGIAPFFLKPTWKDILIN 125
Db 87 IFTAKDEAKDQV-ALLAQVGSINIHESIEIGKVATELGYNCLSAVTPFYK-TFFPEIK 144
QY 126 NPLKEVAAAAPALPFYYHYHIPALTGVKIRAEELLDGILKIPFQGLKFSDDTLDD 185
Db 145 NYITIVNAT-GMNIVYVIPALTGVSWTADQF--GELFENPKIIGVFTAGDFYLLERV 201
QY 186 VDQNRQOQAFELGVDEQLLSALVWGATGAVS 218
Db 202 --KRAYPDHLIWAQFDEMMLPACSLIGIDGAGS 232

RESULT 14
Q8E7V7 PRELIMINARY; PRT; 305 AA.
AC Q8E7V7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN GBS0038.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=NEM316 / Serotype III;
RX MEDLINE=2242528; PubMed=12354221;
RA Glaser P., Rusznick C., Buchrieser C., Chevalier F., Frangeul L.,

```

RA Msadek T., Zouine M., Couve E., Lalicic L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766843; CAD45683.1; -;
DR SAGALIST; gbs0038; -;
DR InterPro; IPR002220; DHDPS.
DR InterPro; IPR000437; Prok_lipoprot_s.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PRO0146; DHPICSNTHASE.
DR ProDom; PD001859; LHDPS; 1.
DR ProSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 305 AA; 33495 MW; 89E06287CAC18C15 CRC64;

Query Match 21.3%; Score 247; DB 16; Length 305;
Best Local Similarity 32.1%; Pred. No. 7.4e-13;
Matches 68; Conservative 40; Mismatches 98; Indels 6; Gaps 6;

QY 6 XKQGLVAACITPMTENGESINFSVIGQYVDYLVKEQGVKNIFVNGTGTGEGLSLSVSERRQ 65
DB 5 QKVGIIIPAFYACYDDKGDICPERVKALTNYSF-DKGVOGLYVNGSSGECIYQSVADKRL 63

QY 66 VAEWVTGKDKLDQVTHVHGALSLESQELACHAAEIGADGIVTAPFELKPTWTDILI 125
DB 64 VLENVMSVAKGKL-TVIAHVACNNTKDSVELAMHAEIGVDAIAALPTIYFR-LPEYAL 121

QY 126 NFLKEVAAAPALPFYVYHIPALTVGKIRAEELDGLDKIPTFQGLKFSDDTLDDFGQC 185
DB 122 DYNTISQAAPQDTFIYINIPQLAGVAL-TSDLYRKMQLN-PQVIGVKNSSMPVQIQNF 179

QY 186 VQNRQQQFAFLFGVDEQLLSALVMGATGAVG 217
DB 180 VAIGGENHIVF-NGPDEQFLGRLMGAAAGIG 210

RESULT 15
Q8E2E9 PRELIMINARY; PRT; 305 AA.
AC Q8E2E9;
DT 01-MAR-2003 (TrEMBLrel. 21, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-acetylneuraminate lyase, putative.
GN SAG0039.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,
Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
Rinaudo D., Rappuoli R., Telford J.B., Kasper D.B., Grandi G.,
Fraser C.M.;
RA "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR EMBL; AE014193; AAM98947.1; -;
DR TIGR; SAG0039; -;
DR GO; GO:0016829; F:lyase activity; IEA.
DR InterPro; IPR002220; DHDPS.
DR InterPro; IPR000437; Prok_lipoprot_s.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PRO0146; DHPICSNTHASE.

DR ProDom; PD001859; DHDPS; 1.
DR ProSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW lyase; Complete proteome.
SQ SEQUENCE 305 AA; 33479 MW; 89E062907C1AE615 CRC64;

Query Match 21.3%; Score 247; DB 16; Length 305;
Best Local Similarity 32.1%; Pred. No. 7.4e-13;
Matches 68; Conservative 40; Mismatches 98; Indels 6; Gaps 6;

QY 6 XKQGLVAACITPMTENGESINFSVIGQYVDYLVKEQGVKNIFVNGTGTGEGLSLSVSERRQ 65
DB 5 QKVGIIIPAFYACYDDKGDICPERVKALTNYSF-DKGVOGLYVNGSSGECIYQSVADKRL 63

QY 66 VAEWVTGKDKLDQVTHVHGALSLESQELACHAAEIGADGIVTAPFELKPTWTDILI 125
DB 64 VLENVMSVAKGKL-TVIAHVACNNTKDSVELAMHAEIGVDAIAALPTIYFR-LPEYAL 121

QY 126 NFLKEVAAAPALPFYVYHIPALTVGKIRAEELDGLDKIPTFQGLKFSDDTLDDFGQC 185
DB 122 DYNTISQAAPQDTFIYINIPQLAGVAL-TSDLYRKMQLN-PQVIGVKNSSMPVQIQNF 179

QY 186 VQNRQQQFAFLFGVDEQLLSALVMGATGAVG 217
DB 180 VAIGGENHIVF-NGPDEQFLGRLMGAAAGIG 210

Search completed: May 6, 2004, 09:07:28
Job time : 26.7628 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:59:54 ; Search time 9.57633 Seconds
(without alignments)
1239.929 Million cell updates/sec

Title: US-09-930-440B-2
Perfect score: 1.59
Sequence: 1 MAFFPKKLGVAATITPTM.....GATGAVGSEVSRDLSTLLSN 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2.6/prodata/2/iaa/5A COMB.pap.*
- 2: /cgn2.6/prodata/2/iaa/5B COMB.pap.*
- 3: /cgn2.6/prodata/2/iaa/6A COMB.pap.*
- 4: /cgn2.6/prodata/2/iaa/6B COMB.pap.*
- 5: /cgn2.6/prodata/2/iaa/ECTUS COMB.pap.*
- 6: /cgn2.6/prodata/2/iaa/backfiles1.pap.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1159	100.0	230	US-09-516-143A-6	Sequence 6, Appli
2	1104.5	95.3	320	US-09-027-013-1	Sequence 1, Appli
3	1104.5	95.3	320	US-09-244-233-1	Sequence 1, Appli
4	394	34.0	300	US-09-543-681A-6868	Sequence 6868, Ap
5	276.5	23.9	297	US-09-027-013-3	Sequence 3, Appli
6	276.5	23.9	297	US-09-244-233-3	Sequence 3, Appli
7	269.5	23.3	308	US-09-543-681A-4426	Sequence 426, Ap
8	248.5	21.4	318	US-09-027-013-4	Sequence 4, Appli
9	248.5	21.4	318	US-09-244-233-4	Sequence 4, Appli
10	241.5	20.8	341	US-09-328-352-5092	Sequence 5092, Ap
11	239.5	20.7	302	US-09-107-532A-5280	Sequence 5280, Ap
12	222.5	19.2	301	US-08-674-168-31	Sequence 31, Appl
13	222.5	19.2	301	US-08-985-908-15	Sequence 15, Appl
14	222.5	19.2	301	US-08-852-730-20	Sequence 20, Appl
15	220	19.0	294	US-09-489-039A-9498	Sequence 9498, Ap
16	214	18.5	290	US-09-631-828A-2	Sequence 2, Appli
17	207.5	17.9	299	US-09-134-001C-5221	Sequence 5221, Ap
18	196.5	17.0	299	US-09-252-991A-29593	Sequence 1976, Ap
19	196.5	17.0	318	US-09-540-326-1976	Sequence 5386, Ap
20	177	15.3	297	US-08-328-352-5386	Sequence 4, Appli
21	170	14.7	292	US-08-648-010-4	Sequence 4, Appli
22	170	14.7	292	US-08-973-461A-4	Sequence 5174, Ap
23	169.5	14.6	301	US-09-134-000C-5174	Sequence 9165, Ap
24	165	14.2	295	US-09-489-039A-9165	Sequence 20599, A
25	162	14.0	334	US-09-252-991A-20599	Sequence 111, App
26	158	13.6	253	US-09-724-623-111	Sequence 6781, Ap
27	158	13.6	314	US-09-328-352-6781	

ALIGNMENTS

RESULT 1

US-09-516-143A-6
; Sequence 6, Application US/09516143A
; Patent No. 6333182
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PFS05PCT
; CURRENT APPLICATION NUMBER: US/09/516,143A
; CURRENT FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-516-143A-6

Query Match 100.0%; Score 1159; DB 4; Length 230;

Best Local Similarity 100.0%; Pred No. 2.5e-125;

Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAFPKKKLGVAATITPTMTENGEINFSV-GQYVDYLKQGVKNIFVNGTTGGLSV 60	
DB	1	MAFPKKKLGVAATITPTMTENGEINFSV-GQYVDYLKQGVKNIFVNGTTGGLSV 60	
QY	61	SRRQVAEEVWTGKDKLDQVIHVGALSLSKESQELACHAAEIGADGIAVIAFPFLKEWT 120	
DB	61	SRRQVAEEVWTGKDKLDQVIHVGALSLSKESQELACHAAEIGADGIAVIAFPFLKEWT 120	
QY	121	KDILINFLKVAAPALPFVYHHPALTGKIPABELLGILDKIPTFQGLKFSDDL 180	
DB	121	KDILINFLKVAAPALPFVYHHPALTGKIPABELLGILDKIPTFQGLKFSDDL 180	
QY	181	DFGQCVQNRQQQFAFLFGVDEQLLSALWMGATGAVGSEVSRDLSTLLSN 230	
DB	181	DFGQCVQNRQQQFAFLFGVDEQLLSALWMGATGAVGSEVSRDLSTLLSN 230	

RESULT 2

US-09-027-013-1
; Sequence 1, Application US/09027013
; Patent No. 5962302
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE

CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/05/344,233
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 09/027,013
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0462 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 320 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT07
CLONE: 2125668
US-09-244-233-1

Query Match	95.3%	Score 1104.5	DB 3	Length 320
Best Local Similarity	91.0%	Pred. No. 8.1e-119		
Matches 222	Conservative	2	Mismatches 5	Indels 15 Gaps 1
QY	1	MAFPKKKLGGLVAAATTPMTENGEINFSVIGQVQVYDYLWKQGVKNIFVNGTTGEGLSLSV	60	
Db	1	MAFPKKKLGGLVAAATTPMTENGEINFSVIGQVQVYDYLWKQGVKNIFVNGTTGEGLSLSV	60	
QY	61	SRRQVAEHWVTKGDKLDQVITHVGLSLSKESQELAQHAAEIGADGIAVIAPFLAPWPT	120	
Db	61	SRRQVAEHWVTKGDKLDQVITHVGLSLSKESQELAQHAAEIGADGIAVIAPFLAPWPT	120	
QY	121	KDILINFLKEVAAAAAPALPFYVYHIFALTVGKIRABEELGILDIKITPFGQLKESDITLL	180	
Db	121	KDILINFLKEVAAAAAPALPFYVYHIFALTVGKIRABEELGILDIKITPFGQLKESDITLL	180	
QY	181	DFGCGQVDNQEQOFALFPGVDEQLLSALVMGATGAVGS-----FVSRDLS	225	
Db	181	DFGCGQVDNQEQOFALFPGVDEQLLSALVMGATGAVGSTNYLKGKTNQMLEAFEQKDF	240	
QY	226	TLIS	229	
Db	241	IAIN	244	

RESULT 4
US-09-543-681A-6868
Sequence 6868, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETTON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543.681A
CURRENT FILING DATE: 2000-04-05

Thu May 13 11:52:59 2004

```

; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6868
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6868

Query Match
Best Local Similarity 34.0%; Score 394; DB 4; Length 300;
Matches 87; Conservative 39; Mismatches 82; Indels 4; Gaps 4;

QY 7 KLGGLVAATITPMTENGESINFSVIGQVYVLYVKEQGVKNIFVNGTTGEGLSLSVSERRQV 66
Db 6 KLSGLIAAHTFPAADGSVNPVIDDIAKHLI-TTGTGAYVLGTTGEGHCSVEERKXV 64
QY 67 AEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFFLKPTWKDILIN 126
Db 65 AERWVTASQGLD-LIHTGALSIADTLEARHAETLDIKATSVIGPCFFKPSHVDLVE 123
QY 127 ELKEVAAAAPALPFYVYHPIALTCVKIRAEELDGILDKIPTFQGLKFSDDLDFGQCV 186
Db 124 YCLAAASAPSKGFYYH-STWSGLSDMEKFLQAAAGKVLENLGGKFNPSFDMYEFQRL 182
QY 187 DQNRQOQFAFLF-GVDEQLLSALVNGATGAVGS 218
Db 183 -RVEGKYLIPFGVDFEIPAGLACGALSAGVS 213
```

```

RESULT 5
US-09-027-013-3
; Sequence 3, Application US/09027013
; Patent No. 5962302
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/027,013
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0462 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: GI 42131
US-09-244-233-3
```

```

; LIBRARY: GenBank
; CLONE: GI 42131
US-09-027-013-3

Query Match
Best Local Similarity 23.9%; Score 276.5; DB 2; Length 297;
Matches 59; Conservative 48; Mismatches 86; Indels 9; Gaps 6;

QY 8 LQGLVAATITPMTENGESINFSVIGQVYVLYVKEQGVKNIFVNGTTGEGLSLSVSERRQVA 67
Db 5 LRGVMAALLTPDQOQALDKASLRRLVQFNI-QQGIIDGLYVGGSTGEAFVQSLSEREQVL 63
QY 68 EEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFFLKPTWKDILIN 127
Db 64 EIVAEKGGKI-KLIAHVGCVTITRESQQLAASAKRYGFDASVATPPYY-PFSEEHCDH 121
QY 128 LKEVAAAAPALPFYVYHPIALTCVKIRAEELDGILDKIPTFQGLKFSDDLDFGQCV 187
Db 122 YRAIIDSADGLPMVYVNIIPALSGVLTLDQI--NTLVTLPVGALKQTSGLDYQMEQI-- 177
QY 188 QNRQOQFAFLF-GVDEQLLSALVNGATGAVGS 218
Db 178 -RREHDELVLYNGYDSEIFASGLLAGADGGIGS 208
```

```

RESULT 6
US-09-244-233-3
; Sequence 3, Application US/09244233
; Patent No. 6030824
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,233
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/027,013
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0462 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: GI 42131
US-09-244-233-3

Query Match
23.9%; Score 276.5; DB 3; Length 297;
```


STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA: US/09/027,013
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0462 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: GI 1016806
US-09-027-013-4

Query Match 21.4%; Score 248.5; DB 2; Length 318;
Best Local Similarity 30.0%; Pred. No. 4.4e-20;
Matches 64; Conservative 47; Mismatches 95; Indels 7; Gaps 5;

QY 6 KKLQGLVAATIPMTENGESVIGQYVDYLVKQGVKNIFVNGTGGSLSVSRQ 65
DB 27 KSLKGLFSAALVSFNEGTINEKGLREIVRNIDKIDGLVGGSTGENFELSTEKKQ 86
QY 66 VAEWVTYKSKDQVLIHVGLSKESQELAQHAEIGADGIAVIAPFLKPTWKDILI 125
DB 87 IFRIAKDEAKQV-ALLAQVGSINIHESITELGKATYELGYNCLSAVTPFYK-FTPEIK 144
QY 126 NFLKEVAAAPALPFPYVYHIPALTVKIRAEILLDGLDKIPTFQGLKFSDTLLDFGQC 185
DB 145 NYNTIVNAT-GMMIVVSIPTALTVGSMTRDQF--GELFENKILGVKFTAGDFYLLERV 201
QY 186 VQNRQQQFAFLFGVDEQLLSALVMGATGAVS 218
DB 202 --KRAYPHLIWAGFEDENMLPACSLGIDGAIGS 232

RESULT 9
US-09-244-233-4
Sequence 4, Application US/09244233
Patent No. 6030824
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

Best Local Similarity 32.5%; Pred. No. 2.3e-23;
Matches 69; Conservative 48; Mismatches 86; Indels 9; Gaps 6;

QY 8 LQGLVAATIPMTENGESVIGQYVDYLVKQGVKNIFVNGTGGSLSVSRQVA 67
DB 5 LRGVMAALLPFFQQQLDKASRLVQFNI-QGIDGLVGGSTGAFVQSLSERQVL 63
QY 68 EEWVTYKSKDQVLIHVGLSKESQELAQHAEIGADGIAVIAPFLKPTWKDILINF 127
DB 64 EIVAEEGKGI-KLIAHVGCVTTAESQOLAASAKRYGFDVAIVTPFY-PPSPSEHCDE 121
QY 128 LKEVAAAPALPFPYVYHIPALTVKIRAEILLDGLDKIPTFQGLKFSDTLLDFGQCD 187
DB 122 YRAIIDSADGLPMVYNIPALSGVKLTDQI--NTLVTLPQGVGALKQTSGLDYQMEI-- 177
QY 188 QNRQQQFAFLF-GVDEQLLSALVMGATGAVS 218
DB 178 -RREHPLVLVNGYDEIFASGLLAGADGGIGS 208

RESULT 7
US-09-543-681A-4426
Sequence 4426, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 4426
LENGTH: 308
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-4426

Query Match 23.3%; Score 269.5; DB 4; Length 308;
Best Local Similarity 34.9%; Pred. No. 1.6e-22;
Matches 75; Conservative 34; Mismatches 93; Indels 13; Gaps 7;

QY 6 KKLQGLVAATIPMTENGESVIGQYVDYLVKQGVKNIFVNGTGGSLSVSRQ 65
DB 7 KKLFGVTAMVTPFDSNCHVDINALSALTNNLI-TKGVDCLYPCGHTGEMLSASERKS 65
QY 66 VAEWVTYKSKDQVLIHVGLSKESQELAQHAEIGADGIAVIAPFLKPTWKDILI 125
DB 66 VAQTVYETANKRL-PVFIHVGMATVTEITELAKHAEIGADGIGVTPPQFFAATDRE--LE 123
QY 126 NFLKEVAAAPALPFPYVYHIPALTVKIRAEILLDGLDKIPTFQGLKFSPTD---LLD 181
DB 124 HYFVIVANSVPENFPYLYNIQCAANNIK-PELAAKIQQCKNIGIKYSPADNNTTIG 182
QY 182 FGQCVQNRQQQFAFLFGVDEQLLSALVMGATGAV 216
DB 183 YLAVAD-----NFSVLHGYDKLFLGLLDAGCDGTV 212

RESULT 8
US-09-027-013-4
Sequence 4, Application US/09027013
Patent No. 5962302
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,233
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 39/027,013
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0462 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: GI 1016806
US-09-244-233-4

Query Match 21.4%; Score 248.5; DB 3; Length 318;
Best Local Similarity 30.0%; Pred. No. 4.4e-20;
Matches 64; Conservative 47; Mismatches 95; Indels 7; Gaps 5;

QY 6 KXQLGLVAATITPTMTENGINSVIGQYVDYLVYKEQGVKNIFVNGTTGEGLSLSVSRQYA 65
DB 27 KXSLGLFSALLVSFNEGDTINEKGLREIVRYNIDKMKIDGLYVGGSTGENFELSTEEKQ 86
QY 66 VAEWVTKGKDKLQVLIHVGALSLSKESQELAHAAIGADGIAVIAPFFFLKPKWTKDILI 125
DB 87 IFRIAKDEAKQV-ALIAQVGSINIHESIELGLYATELGNCLSAVTPPYK-FTFEIIX 144
QY 126 NELKEVAAAPALPFYYPHYIPALTGVKIRAEELDGLDKIPTFQGLKFSDDTLDDFGQC 185
DB 145 NYNIVNAT-GNMIVYISIPALTGVSMADQF-GELEPNKIIIGVKTFAGFYLLERV 201
QY 186 VQNTQQQAFPLFGVDEQLLSALVMGATGAVS 218
DB 202 --KRAYPDHLIWAAGFDEMMLPACSLGIDGAIGS 232

RESULT 10
US-09-328-352-5092
Sequence 5092, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5092
LENGTH: 341
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5092

Query Match 20.8%; Score 241.5; DB 4; Length 341;
Best Local Similarity 30.6%; Pred. No. 3.1e-19;
Matches 64; Conservative 49; Mismatches 89; Indels 7; Gaps 6;

QY 8 LQGLVAATITPTMTENGINSVIGQYVDYLVYKEQGVKNIFVNGTTGEGLSLSVSRQYA 67
DB 52 TQGSIVATVTPMLKDGVDWKSLEKLVEMHI-EQGTNSIVAVGTTGEGSTLSMEETQVI 110
QY 68 BEWVTKGKDKLQVLIHVGALSLSKESQELAHAAIGADGIAVIAPFFFLKPKWTKDILI 127
DB 111 KE-IIRVANKRIPILANGGANSTRAIELTKAAGADAAALLVTPYINKP-TQEGLYOH 168
QY 128 LKEVAAAPALPFYYPHYIPALTGVKIRAEELDGLDKIPTFQGLKFSDDTLDDFGQCVD 187
DB 169 YKAIARAV-ELFLILYNVGRGTGVDLSNDTAVR-LAEIPNIVGIDKATGD-VPRGKALI 224
QY 188 QNRQQQAFPLFGVDEQLLSALVMGATGAV 216
DB 225 DALNGKMAVYSGDDETAWEMLMLGADGNI 253

RESULT 11
US-09-107-532A-5280
Sequence 5280, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A. Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <UNKNOWN>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,499
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5280:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...302
SEQUENCE DESCRIPTION: SEQ ID NO: 5280:
US-09-107-532A-5280

Query Match 20.7%; Score 239.5; DB 4; Length 302;
Best Local Similarity 29.4%; Pred. No. 4.4e-19;
Matches 63; Conservative 51; Mismatches 87; Indels 13; Gaps 7;

QY 6 KXQLGLVAATITPTMTENGINSVIGQYVDYLVYKEQGVKNIFVNGTTGEGLSLSVSRQ 65

Db 9 KMKIGSIVALTIPMEEGSDVAGLEKLIQFHDEQ--TGCLLVLTGTSSTLTQSEEEQ 67
QY 66 VABEWTKGKDLQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFLKPKWKDILI 125
Db 68 ILQLTVKVAGRV-PVIAGAGTNTKETTIEKAHFAS--GADALVITPYNKTSAGLAA 126
QY 126 NFKUVAAPALPFFVYHHPALTVGKIRAEELDGLDKIPTFQGLKFSDTDL--LDF 182
Db 127 HF-TAARASP-IPLIILYNVPSRTGMSIPTHLVN--LAEHPNIGLKEASGDMAVYMDA 182
QY 193 GQCVDQNRQOQFAPLFGVDEQLLSALVMGATGAV 216
Db 183 ARLIG----EEFFLYSGNDLILPVMVGGGWI 212

RESULT 12
US-08-674-168-31
; Sequence 31, Application US/08674168
; Patent No. 5804414
; GENERAL INFORMATION:
; APPLICANT: MORIYA, Mika
; APPLICANT: MASUI, Hiroshi
; APPLICANT: YOKOZAKI, Kenzo
; APPLICANT: HIRANO, Seiko
; APPLICANT: HAYAKAWA, Atsushi
; APPLICANT: IZUI, Masako
; APPLICANT: SUGIMOTO, Masakazu
; TITLE OF INVENTION: METHOD OF AMPLIFYING GENE USING
; TITLE OF INVENTION: ARTIFICIAL TRANSDUCTION
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P.C.
; ADDRESS: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; STREET: ARLINGTON
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,168
; FILING DATE: 01-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-166541
; FILING DATE: 30-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-674-168-31
Query Match 19.2%; Score 222.5; DB 1; Length 301;
Best Local Similarity 29.6%; Pred. No. 4e-17;
Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;
QY 14 ATITPMTEGEINFSVIGQYVDLYVKEQGVKNIFVNGTTGEGLSLSVSRQVAEHWTK 73

Db 20 AMVTFESGDIDIAAGREVAAYLV-DKGLDSLVLGAGTTGESPPTTAAEKELELLKAVREE 78
QY 74 GKDKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFLKPKWKDILIINFLKXVAA 133
Db 79 VGDRAN-VIAGVTNNTRTSVLEAAASAGADGLLVITPYYSKPSQEGLLAHP--GAIA 135
QY 134 AAPALPFFVYHHPALTVGKIRAEELDGLDKIPTFQGLKFSDTDLDFGQCVQDQNRQOQ 193
Db 136 AATEVPCILYDIPGRSGIPESDTMR--LSELPITLAVKADKGLVAATSLI---KETG 190
QY 194 FAELFGVDEQLLSALVMGATGAV 216
Db 191 LAWVGDDPLNLVWALGSGSFI 213

RESULT 13
US-08-985-908-15
; Sequence 15, Application US/08985908
; Patent No. 6024773
; GENERAL INFORMATION:
; APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSETYOSEI NA
; TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,908
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-325659
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-985-908-15
Query Match 19.2%; Score 222.5; DB 3; Length 301;
Best Local Similarity 29.6%; Pred. No. 4e-17;
Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;
QY 14 ATITPMTEGEINFSVIGQYVDLYVKEQGVKNIFVNGTTGEGLSLSVSRQVAEHWTK 73
Db 20 AMVTFESGDIDIAAGREVAAYLV-DKGLDSLVLGAGTTGESPPTTAAEKELELLKAVREE 78
QY 74 GKDKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFLKPKWKDILIINFLKXVAA 133
Db 79 VGDRAN-VIAGVTNNTRTSVLEAAASAGADGLLVITPYYSKPSQEGLLAHP--GAIA 135
QY 134 AAPALPFFVYHHPALTVGKIRAEELDGLDKIPTFQGLKFSDTDLDFGQCVQDQNRQOQ 193
Db 136 AATEVPCILYDIPGRSGIPESDTMR--LSELPITLAVKADKGLVAATSLI---KETG 190
QY 194 FAELFGVDEQLLSALVMGATGAV 216

us-09-930-440b-2.ra1

Thu May 13 11:52:59 2004

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001

CURRENT APPLICATION NUMBER: US/09/489,039A

CURRENT FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: US 60/117,747

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 9498

LENGTH: 294

TYPE: PRT

ORGANISM: Klebsiella pneumoniae

US-09-489-039A-3498

Query Match 19.0%; Score 220; DB 4; Length 294;

Best Local Similarity 28.8%; Pred. No. 7.4e-17;

Matches 65; Conservative 40; Mismatches 97; Indels 24; Gaps 8;

QY 5 KKKLOGLVAATITMTENGESVIGVGVYDYLKKEQGVKN-IFVNGITGEGLSLSVSR 63

DB 2 RRAISGVLTALVTPPTAEGALNLPALRQVQ---RQLAAGNGIFCGGTNGEFFFVLEEEK 58

QY 64 EQVABEWTGKDKLDQVTHVGAALSLXESQELAQHAAIGADGIAVIAPFLKPTKDI 123

DB 59 LAVARTCVBEAAGRA-PVVAHIGEVSTRETRRLGQOLARLGVDVAISATEFV-PLKQEE 116

QY 124 LINFLEKVAAPALPFYHYHIPALTGVKIRAE-----ELLGILDKIETFOGLKFS 175

DB 117 LINHYTALADAL-SVPEFLYNIPARTGNTIAPETARQARHENIVGIKDSAGSIDSLK-- 173

QY 176 DTDLLDFGQCVQDNRQOQFAFLFGVDEQLLSALVMGATGAVGSFVS 221

DB 174 --GFLDAVRDID-----GFDVLNGPDSLHQGFVDGCSACISGLAN 212

Search completed: May 6, 2004, 09:08:35

Job time: 10.5763 secs

Db 191 LAWYSGDDPLNLVWALGGSGFI 213

RESULT 14
US-08-852-730-20
Sequence 20, Application US/08852730

Patent No. 6096597

GENERAL INFORMATION:

APPLICANT: SEIKO HIRANO, MASAKAZU SUGIMOTO, EICHI NAKANO,

APPLICANT: MASAKO IZUT, ATSUSHI HAYAKAWA, YASUHIKO YOSHIHARA, AND TSUYOSHI

APPLICANT: NAKAMATSU

TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESSES: 24

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MALER AND NEUSTADT, P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY

CITY: ARLINGTON

STATE: VA

ZIP: 22026

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/852,730

FILING DATE: 05-07-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-142812

FILING DATE: 05-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 301 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLSCULE TYPE: protein

US-08-852-730-20

Query Match 19.2%; Score 222.5; DB 3; Length 301;

Best Local Similarity 29.6%; Pred. No. 4e-17;

Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;

QY 14 ATITMTENGESVIGVGVYDYLKKEQGVKNIFVNGITGEGLSLSVSRQVAAEWVTK 73

DB 20 AMVTFTESGDDIDIAAGREVAAYLV-DKGLDSLVLVAGTTGESPTTAAEKLELLKAVREE 78

QY 74 GKDKLDQVTHVGAALSLXESQELAQHAAIGADGIAVIAPFLKPTKDIILNFKVEVA 133

DB 79 VGDRAV-VIAGVGTNNRTSLVELAAASAGADGILLVTPYYSKPSQEGGLAHP--GAIA 135

QY 134 AAPALPFYTHIPALTGVKIRAEELDGLDKPTFGGLKFSDDTLDDFCQCVQDNRQOQ 193

DB 136 AATEVPICLYDIPRSGIPESDTMER--LSELPITLAVKADKGLVAATSLLI---KETG 190

QY 194 FAFLEFGVDEQLLSALVMGATGAV 216

DB 191 LAWYSGDDPLNLVWALGGSGFI 213

RESULT 15

US-09-489-039A-9498

Sequence 9498, Application US/09489039A

Patent No. 6610836

GENERAL INFORMATION:

APPLICANT: Gary Breton et. al

Sequence 85, Appli
Sequence 45588, A
Sequence 16546, A
Sequence 43119, A
Sequence 23683, A
Sequence 13888, A
Sequence 56432, A
Sequence 716, Appli
Sequence 2, Appli
Sequence 21028, A
Sequence 55465, A
Sequence 19025, A
Sequence 18247, A
Sequence 14458, A
Sequence 14914, A
Sequence 11732, A
Sequence 17315, A
Sequence 23132, A
Sequence 336, Appl
Sequence 20676, A
Sequence 21450, A
Sequence 4687, Ap
Sequence 55, Appli
Sequence 2859, Ap
Sequence 46924, A
Sequence 7446, Ap
Sequence 19626, A
Sequence 17250, A
Sequence 9753, Ap
Sequence 20274, A

15 US-10-369-493-85
12 US-10-282-122A-45588
15 US-10-369-493-16546
12 US-10-282-122A-43119
15 US-10-369-493-23683
12 US-10-369-493-13888
15 US-10-282-122A-56432
12 US-10-369-493-716
15 US-10-214-556-2
12 US-10-369-493-21028
15 US-10-282-122A-55465
12 US-10-369-493-19025
15 US-10-369-493-18247
12 US-10-369-493-14458
15 US-10-369-493-14914
12 US-10-369-493-11732
15 US-10-238-075-336
12 US-10-369-493-23132
15 US-10-369-493-20676
12 US-10-369-493-21450
15 US-10-369-493-4687
9 US-09-971-536-55
15 US-10-369-493-2859
12 US-10-282-122A-46924
15 US-10-369-493-7446
12 US-10-369-493-19626
15 US-10-369-493-17250
12 US-10-369-493-9753
15 US-10-369-493-20274

222 19.2 294 15
219 18.9 298 12
218.5 18.9 289 15
218.5 18.9 319 12
218.5 18.9 319 15
218 18.8 291 15
218 18.8 309 12
218 18.8 309 15
218 18.8 309 15
214 18.5 290 14
214 18.5 302 15
210 18.1 292 12
208 17.9 291 15
205.5 17.7 292 15
205.5 17.7 293 15
205.5 17.7 293 15
204.5 17.6 307 15
204.5 17.6 292 15
203.5 17.6 295 14
203.5 17.6 286 15
201.5 17.4 289 15
201 17.3 290 15
201 17.3 301 9
201 17.3 301 15
199.5 17.2 291 12
198 17.1 288 15
197.5 17.0 290 15
196.5 17.0 295 15
196 16.9 289 15
196 16.9 291 15

US-09-930-440B-2
1159
1 MAFPKKKQLGLVAATITPMT.....GATGAVGSFVRDLSTLLSN 230
BLOSUM62
Gapop 10.0 , Gapext 0.5
1140673 seqs, 277566755 residues
1140673
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*

ALIGNMENTS

RESULT 1

US-09-984-205-6
; Sequence 6, Application US/09984205
; Patent No. US20020137175A1
; GENERAL INFORMATION:
; APPLICANT: Coleman, Timothy A. et al.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: P5505D1
; CURRENT APPLICATION NUMBER: US/09/984,205
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: PCT/US00/05325
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/516,143
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-205-6

Query Match 100.0%; Score 1159; DB 9; Length 230;
Best Local Similarity 100.0%; Pred. No. 8.5e-119;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFPKKKQLGLVAATITPMTENGEINFSVIGQYDYLKKEQGVKIFVNGTTEGLSLSV 60
Db 1 MAFPKKKQLGLVAATITPMTENGEINFSVIGQYDYLKKEQGVKIFVNGTTEGLSLSV 60

QY 61 SERROVAEWYTKGKLDQVLIHVGALSLKESQSLAAEIGADGIAVIAPFFPKWT 120
Db 61 SERROVAEWYTKGKLDQVLIHVGALSLKESQSLAAEIGADGIAVIAPFFPKWT 120

QY 121 KDILNFKVAAPALPFVYVHIPALTGKIRAEELLDGILDKIPTFGKFSFDLL 180

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1159	100.0	230	9	US-09-984-205-6
2	1159	100.0	230	9	US-09-930-440B-2
3	274.5	23.7	297	15	US-10-369-493-23570
4	272.5	23.5	297	14	US-10-182-504-14
5	251.5	21.7	294	15	US-10-369-493-21648
6	241	20.8	313	15	US-10-369-493-18552
7	234.5	20.2	299	15	US-10-369-493-8308
8	231.5	20.0	283	15	US-10-369-493-1171
9	230.5	19.9	287	15	US-10-369-493-1265
10	230.5	19.9	287	15	US-10-369-493-20331
11	226	19.5	290	15	US-10-369-493-11339
12	222.5	19.2	301	7	US-08-952-976-15
13	222.5	19.2	301	9	US-09-738-626-6976
14	222.5	19.2	301	10	US-09-746-660A-54
15	222.5	19.2	301	14	US-10-226-136-15

121 KDILNFKLKEVAAAPALPFFYYHIPALTVGKIRABELLDGILDKIPTFQGLKFSDDTL 180
181 DFGQCVQNRQQQFAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLSN 230
181 DFGQCVQNRQQQFAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLSN 230

RESULT 2

US-09-930-440B-2
Sequence 2, Application US/09930440B
Patent No. US20020142386A1
GENERAL INFORMATION:
APPLICANT: Setenbaugh et al.
TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
FILE REFERENCE: PFS0982
CURRENT APPLICATION NUMBER: US/09/930,440B
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 60/227,579
PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: 09/516,793
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/169,624
PRIOR FILING DATE: 1999-12-08
PRIOR APPLICATION NUMBER: 60/122,582
PRIOR FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 230
TYPE: PRT
ORGANISM: Homo sapiens
US-09-930-440B-2

Query Match 100.0%; Score 1159; DB 9; Length 230;
Best Local Similarity 100.0%; Pred. No. 8.5e-119;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAFPKKXQLGLVAATITPTMTENGINFVIGQYDYLKKEQVKNIFVNGTTGEGLSLV 60
1 MAFPKKXQLGLVAATITPTMTENGINFVIGQYDYLKKEQVKNIFVNGTTGEGLSLV 60
61 SRRQVAEEWTKGDKLDQVLIHVGALSLSKESQELAQAAEIGADGIAVIAPFFLKPT 120
61 SRRQVAEEWTKGDKLDQVLIHVGALSLSKESQELAQAAEIGADGIAVIAPFFLKPT 120
121 KDILNFKLKEVAAAPALPFFYYHIPALTVGKIRABELLDGILDKIPTFQGLKFSDDTL 180
121 KDILNFKLKEVAAAPALPFFYYHIPALTVGKIRABELLDGILDKIPTFQGLKFSDDTL 180
181 DFGQCVQNRQQQFAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLSN 230
181 DFGQCVQNRQQQFAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLSN 230

RESULT 3

US-10-369-493-23570
Sequence 23570, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 23570

LENGTH: 297
TYPE: PRT
ORGANISM: Escherichia coli
US-10-369-493-23570
Query Match 23.7%; Score 274.5; DB 15; Length 297;
Best Local Similarity 34.3%; Pred. No. 2.1e-21;
Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;

8 LOGLVAAATITPTMTENGINFVIGQYDYLKKEQVKNIFVNGTTGEGLSLSVSRQ-- 65
5 LRGVMAALLTPDQQAALDKASRLRLVQNT--QQGIDGLYVGGSGEAFVQSLSREQL 63
66 --VAEEWTKGDKLDQVLIHVGALSLSKESQELAQAAEIGADGIAVIAPFFLKPT 123
64 EIVAE--ARKGKIL--IAHVCVSTAESQQLAASAKRYGFDVAVTPFFY-PFSFE 117
124 LINFKEVAAAPALPFFYYHIPALTVGKIRABELLDGILDKIPTFQGLKFSDDTL 183
118 HCDHYRAIIDSADGLPMVYVNIIPALSGVKLTLDQI--NTLVLPGVGALKQTSGLYQME 175
184 QVDONROQQAFLP-GYDEQLLSALVMGATGAVGS 218
176 QI---RREHPDLVLYNGYDEIFASGLLAGADGGIGS 208

RESULT 4

US-10-182-504-14
Sequence 14, Application US/10182504
Publication No. US20030157121A1
GENERAL INFORMATION:
APPLICANT: Chatfield, Steven Neville
TITLE OF INVENTION: Virulence Genes, Proteins, and Their Use
FILE REFERENCE: GJE-98
CURRENT APPLICATION NUMBER: US/10/182,504
CURRENT FILING DATE: 2002-07-29
PRIOR APPLICATION NUMBER: PCT/GB01/00449
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: GB 0023059.9
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: GB 0023060.7
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: GB 0002552.8
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: GB 0002554.4
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: GB 0002555.1
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: GB 0002556.9
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: GB 0002557.7
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: GB 0002558.5
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: GB 0002559.3
PRIOR FILING DATE: 2000-02-03
Remaining prior Application data removed - see File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 297
TYPE: PRT
ORGANISM: Salmonella typhimurium
US-10-182-504-14

Query Match 23.5%; Score 272.5; DB 14; Length 297;
Best Local Similarity 31.9%; Pred. No. 3.6e-21;
Matches 68; Conservative 43; Mismatches 95; Indels 7; Gaps 5;

6 KKLQGLVAATITPTMTENGINFVIGQYDYLKKEQVKNIFVNGTTGEGLSLSVSRQ 65
3 KALQGVMAALLTPDQQAALDKASRLRLVQNT--QQGIDGLYVGGSGEAFVQSLAEREQ 61

QY	66	VABEWTKGDKLQDVIHVGALSKLBSQELAAHAE	GGAGIVIAFFKPKPTKCLIL	125
Db	62	VLEIVAEAGKGI-TLAAHGTGTATSSQQLA	AKARYGFDVAVSATPPY-PPSFEBHC	119
QY	126	NFLKEVAARAPALPFYYHYHIALPGWIKARE	LIDGLDKIPTGGLKFSQETLLDFGQC	185
Db	120	DHYRAIIDSADGLPMVYVNPALSGVKLTQOI	-NTLVTLPGVSALKQTSGLDFQMEQI	177
QY	186	VDQNRQQCFAPLFGVDECLLSALVWGATG	AVGS	218
Db	178	--RAHPDULVINGYDEIFASGLLAGADGG	IGS	208

RESULT 5
 US-10-369-493-21648
 ; Sequence 21648, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052) B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 21648
 ; LENGTH: 294
 ; TYPE: PRT
 ; ORGANISM: Pyrococcus abyssi
 US-10-369-493-21648

RESULT 6
US-10-369-493-18552
; Sequence 18552, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

```

; PRIOR APPLICATION NUMBER: US 60/360,039
;
; PRIOR FILING DATE: 2002-02-21
;
; NUMBER OF SEQ ID NOS: 47374
;
; SEQ ID NO 18552
;
; LENGTH: 313
;
; TYPE: PRT
;
; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-18552

```

RESULT 7
 US-10-369-493-8308
 ; Sequence 8308, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 8308
 ; LENGTH: 299
 ; TYPE: PRT
 ; ORGANISM: Thermobifida fusca
 US-10-369-493-8308

Query Match	20.2%	Score	234.5	DB	15	Length	299
Best Local Similarity	23.8%	Pred. No.	5.4e-17				
Matches	68	Conservative	39	Mismatches	110	Indels	11
Gaps	5						
QY	4	PKKKLQGLVAATITPWTENGSEINFSVIGQVDYLVKEQGVKNIFVNGTGTGEEGLSLSYSER	63				
Db	7	FNAPFGQMLTAMITPMLDNGEVDYVARLATYLVDBQRNDGLVNGTTGESATTTDEEK	66				
QY	64	RCVABEWTCKKLDQVLIHVGLSLSKESCELAHAAEIGACGZAVIAPFELKPTWKDI	123				
Db	67	ERIILRTVIDAVGDRA--TIVAGAGNDRTHEZELARTAEAGDGLLVITYYKNPPQGL	125				
QY	124	LINFLKVBAAAAPALPFYYHIPALTGVKIRAESLLDGLIKITPFGLKESDTELLDPG	183				
Db	126	LRHF--TAADATGLPIMLYDIPCRGTGTPIDSETIVR--LAEHERIVANKAKDDL---G	178				
QY	184	CCVDQNRQQQAFILGVDQGLLSALVMGATG---AVGSFVSRLDSTILL	228				
Db	179	ASSWWTSTFDLYVSGSDMLNLPILSIGAEFVSVVGHVVGSELHDMI	226				

Thu May 13 11:53:00 2004

```
Matches 61; Conservative 46; Mismatches 73; Indels 51; Gaps 7;

QY 8 LQGLVAATITPMTENGESVIGQYVDYLKQGVKNIFVNGTTGEGLSLSVSRQVA 67
Db 1 MEGVIVPLITPFKEDHSIDFEALEWLIKFL-BEKGVHGIFINSTGTGFTSLSEERKILA 59
QY 68 BEWVTGKDKLDQV-IHVGALSLSKESQELAHAAEIGADGIAVIAPFFLKPTWKDILIN 126
Db 60 E-----KGEVTSRTYLVGTGTSTFEVIELTKHAKELGADGVIVSPYCYLKEKDAIFKH 115
QY 127 FLKEVAAAAPALPFYYHHPA-----LTGVKIRAEEL--LDGILDK 165
Db 116 F--SWAEKVDIPILYAIIPSCANPIS-LEVVVKLALEYSNVIGVKSVDLSLTLSGLIE- 172
QY 166 IPTFQGLKESDLDLDFGQCVQDNQKQFAPLFGVDEQLLSALVMGATGAV 216
Db 173 -----IKERKDFKVTGLDQVFLPNLILGGGGI 202

RESULT 10
US-10-369-493-20331
; Sequence 20331, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1171
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Methanobacterium thermoautotrophicum
US-10-369-493-1171

Query Match 20.4%; Score 231.5; DB 15; Length 283;
Best Local Similarity 27.3%; Pred. No. 1.1e-16;
Matches 62; Conservative 52; Mismatches 90; Indels 23; Gaps 7;

QY 7 KLOGLVAATITPMTENGESVIGQYVDYLKQGVKNIFVNGTTGEGLSLSVSRQV 66
Db 2 KIECTVWMTFTEDDDVDEAGLRNINYL-ENGVDGLVAGTGTGSA-IITHEQGRM 60
QY 67 ABEWVTGKDKLDQV-IHVGALSLSKESQELAHAAEIGADGIAVIAPFFLKPTWKDILIN 126
Db 61 IDILLDEVNGRV-RWAGAGNSREAGMLVEYAEADAGDAALVITPYNKPQEGLEH 119
QY 127 F--LKEVAAAAPALPFYYHHPALTVKIRAEEL-----LDGIL---DKIPTQGLKPSD 176
Db 120 YTMLEEA-----DPLIYINVSRTGTDIDVDTVAELAKDLGIIGKEASEPDLKVSMLR 175
QY 177 TDLLDFGQCVQDNQKQFAPLFGVDEQLLSALVMGATGAVGVSFVRD 223
Db 176 SRLMDLG-----LDDFTVLSGNDMLTLPMTSMGAEGVISVANVD 215

RESULT 9
US-10-369-493-1265
; Sequence 1265, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1265
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1265

Query Match 19.9%; Score 230.5; DB 15; Length 287;
Best Local Similarity 26.4%; Pred. No. 1.4e-16;
Matches 61; Conservative 46; Mismatches 73; Indels 51; Gaps 7;

QY 8 LQGLVAATITPMTENGESVIGQYVDYLKQGVKNIFVNGTTGEGLSLSVSRQVA 67
Db 1 MEGVIVPLITPFKEDHSIDFEALEWLIKFL-BEKGVHGIFINSTGTGFTSLSEERKILA 59
QY 68 BEWVTGKDKLDQV-IHVGALSLSKESQELAHAAEIGADGIAVIAPFFLKPTWKDILIN 126
Db 60 E-----KGEVTSRTYLVGTGTSTFEVIELTKHAKELGADGVIVSPYCYLKEKDAIFKH 115
QY 127 FLKEVAAAAPALPFYYHHPA-----LTGVKIRAEEL--LDGILDK 165
Db 116 F--SWAEKVDIPILYAIIPSCANPIS-LEVVVKLALEYSNVIGVKSVDLSLTLSGLIE- 172
QY 166 IPTFQGLKESDLDLDFGQCVQDNQKQFAPLFGVDEQLLSALVMGATGAV 216
Db 173 -----IKERKDFKVTGLDQVFLPNLILGGGGI 202

RESULT 11
US-10-369-493-11339
; Sequence 11339, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
```


us-09-930-440b-2.rapb

Thu May 13 11:53:00 2004

APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 11339
 LENGTH: 290
 TYPE: PRT
 ORGANISM: Methanosarcina mazei
 US-10-369-493-11339

Query Match 19.5%; Score 226; DB 15; Length 290;
 Best Local Similarity 29.3%; Pred. No. 4.5e-16;
 Matches 61; Conservative 45; Mismatches 94; Indels 8; Gaps 4;
 QY 9 QGLVAATTPMTENGGEINFSVIGQYVDYLVKEQGVKNIFVNGTTGEGLSLSVSRQVAE 68
 Db 3 EGAPALITFTFKDRIIDREGICNIEF-VERGGVAGIVPCGTGTGESATLSALEHEVID 61
 QY 69 EWTXGKOKLDVTHVGSLSKESQELAHAAEIGADGIAVIAPFFLPKPTKDIILNPL 128
 Db 62 IAVEGSK--VPVIAGTGSNNTGEALQFTKHAADAGVGVLLISPYNNKPNAGLLTHFK 118
 QY 129 KEVAAPALPYYHHPALTVGKIRAEELDGLDKIPTFQGLKPSDTLLDFGQCVQD 198
 Db 119 K--IAEVDVPMWYNIPBRTGCDMPLEIVE--LAKVENIVGIKESAGNIGKVSQILEN 174
 QY 189 NRQQQFAELFGVDEQLLSALVMGATGAV 216
 Db 175 TIDEDFVVISGEDNLTLPILSVGGQGV 202

RESULT 12
 US-08-952-976-15
 Sequence 15, Application US/08952976
 Publication No. US20020086370A1
 GENERAL INFORMATION:
 APPLICANT: OTSUNA, Seiko
 APPLICANT: SUGIMOTO, Masakazu
 APPLICANT: IZUI, Masako
 APPLICANT: HAYAKAWA, Atsushi
 APPLICANT: NAKANO, Eiichi
 APPLICANT: KOBAYASHI, Masaki
 APPLICANT: YOSHIMURA, Yasuhiko
 APPLICANT: NAKAMATSU, Tsuyoshi
 TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
 NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/952,976
 FILING DATE: 8-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 7-140614
 FILING DATE: 07-JUL-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: NORMAN F. OBLON
 REGISTRATION NUMBER: 24,618

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-413-3000
 TELEFAX: 703-413-2220
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 301 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-952-976-15

Query Match 19.2%; Score 222.5; DB 7; Length 301;
 Best Local Similarity 29.6%; Pred. No. 1.1e-15;
 Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;
 QY 14 ATTPMTENGGEINFSVIGQYVDYLVKEQGVKNIFVNGTTGEGLSLSVSRQVAE 73
 Db 20 AMVTPFTESGDIDIAAGREVAAYLV-DKGLDSLVLGAGTTGESPTTAAAEKLELLKAVREE 78
 QY 74 GKDKLDQVTHVGSLSKESQELAHAAEIGADGIAVIAPFFLPKPTKDIILNPLKEVAA 133
 Db 79 VGRAN-VIAGVTNNTSVLLEAAAGAGDGLLVTPYYSKPSQEGLLAHP--GAIA 135
 QY 134 AAPALPYYHHPALTVGKIRAEELDGLDKIPTFQGLKPSDTLLDFGQCVQDNRQQ 193
 Db 136 ATEVPICLYDIPGRSSIPESDTMR--LSELTILAVKDAKGLVAATSLI---KETG 193
 QY 194 FAPLFGVDEQLLSALVMGATGAV 216
 Db 191 LAWYSGDDPLNLVWLALGGSGFI 213

RESULT 13
 US-09-738-626-6976
 Sequence 6976, Application US/09738626
 Publication No. US20020197635A1
 GENERAL INFORMATION:
 APPLICANT: NAKAGAWA, SATOSHI
 APPLICANT: MIZOGUCHI, HIROSHI
 APPLICANT: ANDO, SEIKO
 APPLICANT: HAYASHI, MIKIRO
 APPLICANT: OCHIAI, KEIKO
 APPLICANT: YOKOI, HARUHIKO
 APPLICANT: TATEISHI, NAKO
 APPLICANT: SENO, AKIHIRO
 APPLICANT: IKEDA, MASATO
 APPLICANT: OZAKI, AKIO
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-125
 CURRENT APPLICATION NUMBER: US/09/738,626
 CURRENT FILING DATE: 2000-12-18
 PRIOR APPLICATION NUMBER: JP 99/377484
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: JP 00/159162
 PRIOR FILING DATE: 2000-04-07
 PRIOR APPLICATION NUMBER: JP 00/280988
 PRIOR FILING DATE: 2000-08-03
 NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: PatentIn ver. 3.0
 SEQ ID NO 6976
 LENGTH: 301
 TYPE: PRT
 ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6976

Query Match 19.2%; Score 222.5; DB 9; Length 301;
 Best Local Similarity 30.0%; Pred. No. 1.1e-15;
 Matches 62; Conservative 47; Mismatches 81; Indels 17; Gaps 6;
 QY 14 ATTPMTENGGEINFSVIGQYVDYLVKEQGVKNIFVNGTTGEGLSLSVSR---RQVAEE 69
 Db 20 AMVTPFTESGDIDIAAGREVAAYLV-DKGLDSLVLGAGTTGESPTTAAAEKLELLKAVREE 78

us-09-930-440b-2.rapb

Thu May 13 11:53:00 2004

```

QY 70 WTKGKDLQVLIHVGALSLKESQELAHQAEIGAGS-AVIAPFFLKPTWKDILINFLK 129
Db 79 VGDRAK-----LIAGVTNNTRTSVLEAAASAGAGGLVVT-PYYSKPSQEGLLAHF-- 131
QY 130 EVAAAPALPFYVYHHPALTVKIRAEELDGLDKIPFQGLKFSDDTLDDFGQCVDCN 189
Db 132 GAIAAATEVPICLDIPGRSGIPESDTRR--LSELPILAVKADKGLVAATSLI--- 186
QY 190 RQQQFAFLFGVDEQLLSALVMGATGAV 216
Db 187 KETGLAWYSGDDPLNLVWLALGSGGFI 213

```

RESULT 14

```

US-09-746-660A-54
; Sequence 54, Application US/0974666CA
; Publication No. US20030049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Habernauer, Gregor
; APPLICANT: Kim, Jun-Woc
; APPLICANT: Lee, Heung-Schick
; APPLICANT: Hwang, Byung-Joon
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS
; FILE REFERENCE: BGI-121CP2
; CURRENT APPLICATION NUMBER: US/09/746,660A
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/606740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/603124
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 54
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum

```

US-09-746-660A-54

```

Query Match 19.2%; Score 222.5; DB 10; Length 301;
Best Local Similarity 30.0%; Pred. No. 1.1e-15;
Matches 62; Conservative 47; Mismatches 81; Indels 17; Gaps 6;

QY 14 ATITPMTENGINFVIGQVYLVKEQGVKNIFVNGTTEGSLSVSR-----RQVARE 69
Db 20 AMVTFTESGDIDIAAGREVAAYLV-DKGLDSLVLGTTGSPPTTAAEKLELLKAVREE 78
QY 70 WTKGKDLQVLIHVGALSLKESQELAHQAEIGAGS-AVIAPFFLKPTWKDILINFLK 129
Db 79 VGDRAK-----LIAGVTNNTRTSVLEAAASAGAGGLVVT-PYYSKPSQEGLLAHF-- 131
QY 130 EVAAAPALPFYVYHHPALTVKIRAEELDGLDKIPFQGLKFSDDTLDDFGQCVDCN 189
Db 132 GAIAAATEVPICLDIPGRSGIPESDTRR--LSELPILAVKADKGLVAATSLI--- 186
QY 190 RQQQFAFLFGVDEQLLSALVMGATGAV 216
Db 187 KETGLAWYSGDDPLNLVWLALGSGGFI 213

```

RESULT 15

```

US-10-226-136-15
; Sequence 15, Application US/10226136
; Publication No. US20030054506A1
; GENERAL INFORMATION:
; APPLICANT: OTSUNA, Seiko
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: IZUI, Masako
; APPLICANT: HAYAKAWA, Atsushi
; APPLICANT: NAKANO, Eiichi
; APPLICANT: KOBAYASHI, Masaki
; APPLICANT: YOSHIHARA, Yasuhiko
; APPLICANT: NAKAMATSU, Tsuyoshi
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/226,136
; FILING DATE: 23-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,976
; FILING DATE: 8-DEC-1997
; APPLICATION NUMBER: JP 7-140614
; FILING DATE: 07-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-10-226-136-15

```

```

Query Match 19.2%; Score 222.5; DB 14; Length 301;
Best Local Similarity 29.6%; Pred. No. 1.1e-15;
Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;

QY 14 ATITPMTENGINFVIGQVYLVKEQGVKNIFVNGTTEGSLSVSR-----RQVARE 73
Db 20 AMVTFTESGDIDIAAGREVAAYLV-DKGLDSLVLGTTGSPPTTAAEKLELLKAVREE 78
QY 74 GKDKLDQVLIHVGALSLKESQELAHQAEIGAGS-AVIAPFFLKPTWKDILINFLK 133
Db 79 VGDRAK-----LIAGVTNNTRTSVLEAAASAGAGGLVVT-PYYSKPSQEGLLAHF-- 135
QY 134 AAPALPFYVYHHPALTVKIRAEELDGLDKIPFQGLKFSDDTLDDFGQCVDCN 193
Db 136 AATAATEVPICLDIPGRSGIPESDTRR--LSELPILAVKADKGLVAATSLI--- 190
QY 194 FAFLFGVDEQLLSALVMGATGAV 216
Db 191 LAWYSGDDPLNLVWLALGSGGFI 213

```

US-10-226-136-15

```

Query Match 19.2%; Score 222.5; DB 14; Length 301;
Best Local Similarity 29.6%; Pred. No. 1.1e-15;
Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;

QY 14 ATITPMTENGINFVIGQVYLVKEQGVKNIFVNGTTEGSLSVSR-----RQVARE 73
Db 20 AMVTFTESGDIDIAAGREVAAYLV-DKGLDSLVLGTTGSPPTTAAEKLELLKAVREE 78
QY 74 GKDKLDQVLIHVGALSLKESQELAHQAEIGAGS-AVIAPFFLKPTWKDILINFLK 133
Db 79 VGDRAK-----LIAGVTNNTRTSVLEAAASAGAGGLVVT-PYYSKPSQEGLLAHF-- 135
QY 134 AAPALPFYVYHHPALTVKIRAEELDGLDKIPFQGLKFSDDTLDDFGQCVDCN 193
Db 136 AATAATEVPICLDIPGRSGIPESDTRR--LSELPILAVKADKGLVAATSLI--- 190
QY 194 FAFLFGVDEQLLSALVMGATGAV 216
Db 191 LAWYSGDDPLNLVWLALGSGGFI 213

```

US-10-226-136-15

```

Query Match 19.2%; Score 222.5; DB 14; Length 301;
Best Local Similarity 29.6%; Pred. No. 1.1e-15;
Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;

QY 14 ATITPMTENGINFVIGQVYLVKEQGVKNIFVNGTTEGSLSVSR-----RQVARE 73
Db 20 AMVTFTESGDIDIAAGREVAAYLV-DKGLDSLVLGTTGSPPTTAAEKLELLKAVREE 78
QY 74 GKDKLDQVLIHVGALSLKESQELAHQAEIGAGS-AVIAPFFLKPTWKDILINFLK 133
Db 79 VGDRAK-----LIAGVTNNTRTSVLEAAASAGAGGLVVT-PYYSKPSQEGLLAHF-- 135
QY 134 AAPALPFYVYHHPALTVKIRAEELDGLDKIPFQGLKFSDDTLDDFGQCVDCN 193
Db 136 AATAATEVPICLDIPGRSGIPESDTRR--LSELPILAVKADKGLVAATSLI--- 190
QY 194 FAFLFGVDEQLLSALVMGATGAV 216
Db 191 LAWYSGDDPLNLVWLALGSGGFI 213

```

US-10-226-136-15

```

Query Match 19.2%; Score 222.5; DB 14; Length 301;
Best Local Similarity 29.6%; Pred. No. 1.1e-15;
Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;

QY 14 ATITPMTENGINFVIGQVYLVKEQGVKNIFVNGTTEGSLSVSR-----RQVARE 73
Db 20 AMVTFTESGDIDIAAGREVAAYLV-DKGLDSLVLGTTGSPPTTAAEKLELLKAVREE 78
QY 74 GKDKLDQVLIHVGALSLKESQELAHQAEIGAGS-AVIAPFFLKPTWKDILINFLK 133
Db 79 VGDRAK-----LIAGVTNNTRTSVLEAAASAGAGGLVVT-PYYSKPSQEGLLAHF-- 135
QY 134 AAPALPFYVYHHPALTVKIRAEELDGLDKIPFQGLKFSDDTLDDFGQCVDCN 193
Db 136 AATAATEVPICLDIPGRSGIPESDTRR--LSELPILAVKADKGLVAATSLI--- 190
QY 194 FAFLFGVDEQLLSALVMGATGAV 216
Db 191 LAWYSGDDPLNLVWLALGSGGFI 213

```

us-09-930-440b-2.rapb

Thu May 13 11:53:00 2004

Search completed: May 6, 2004, 09:25:04
Job time : 25.8649 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model.

Run on: May 11, 2004, 07:15:34 ; Search time 5296.32 Seconds
(without alignments)
10679.608 Million cell updates/sec

Title: US-09-930-440b-3
Perfect score: 1305
Sequence: 1 atggactcggaggagaagg.....ataattcgcacaaatag 1305

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 6
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1303.4	99.9	1305	6	BD270234	BD270234 Human gly
2	1280.4	98.1	1305	9	AF397212	AF397212 Homo sapi
3	1272.4	97.5	1376	9	AF271388	AF271388 Homo sapi
4	1272.4	97.5	1766	9	HSMB94286	AL812375 Homo sapi
5	1272.4	97.5	1770	6	AR338869	AR338869 Sequence
6	1270.8	97.4	1634	6	AX882419	AX882419 Sequence
7	1270.8	97.4	1634	6	BD159714	BD159714 Primer fo
8	1270.8	97.4	1694	9	AK022927	AK022927 Homo sapi
9	1270.8	97.4	1772	6	BD260951	BD260951 Carboxydr
10	1270.8	97.4	1772	6	AX040083	AX040083 Sequence
11	1008.8	77.3	1704	10	BC063776	BC063776 Mus muscu
12	1008.8	77.3	1728	10	BC031500	BC031500 Mus muscu
13	1007.2	77.2	1687	10	MM06215	MM06215 Mus muscu
14	918.4	70.4	1588	9	BC016609	BC016609 Homo sapi
15	636.6	48.8	838	6	AX541147	AX541147 Sequence
16	625	47.9	819	6	AX540862	AX540862 Sequence
17	620.2	47.5	748	6	AX541323	AX541323 Sequence
18	562	43.1	686	6	AX541885	AX541885 Sequence
19	538.2	41.2	622	6	AX868002	AX868002 Sequence
20	538.2	41.2	622	6	BD148064	BD148064 Primer fo
21	399	30.6	889	10	BC004606	BC004606 Mus muscu
22	359	27.5	1468	5	AB027414	AB027414 Oncorhync
23	260	19.9	625	9	HS332638	AJ326388 Homo sapi
24	260	19.9	72001	9	AC092862	AC092862 Homo sapi
25	260	19.9	151221	2	AC093092	AC093092 Homo sapi
26	260	19.9	153201	2	AC092863	AC092863 Homo sapi
27	260	19.9	175371	2	AC092822	AC092822 Homo sapi
28	252	19.3	729	9	HS3330107	AJ333107 Homo sapi
29	215	16.5	203003	2	AC087313	AC087313 Homo sapi
30	184.4	14.1	2136	9	AK124990	AK124990 Homo sapi
31	184.4	14.1	153201	2	AC092863	AC092863 Homo sapi
32	184.4	14.1	156314	9	AC087073	BD153135 Primer fo
33	181.6	13.9	589	6	BD153135	AC093092 Homo sapi
34	181.6	13.9	589	6	AC093092	AC119391 Rattus no
35	177.4	13.6	151221	2	AC119391	AC11592 Rattus no
36	174	13.3	213927	2	AC111592	AC092822 Homo sapi
37	174	13.3	216864	2	AC092822	AC087313 Homo sapi
38	162	12.4	175371	2	AC087313	AX310312 Sequence
39	162	12.4	203003	2	AC087313	AC122483 Mus muscu
40	145.8	11.2	222	6	AX310312	AC139119 Didephis
41	145.6	11.2	185388	10	AC122483	AC139119 Rattus no
42	137.2	10.5	201194	2	AC139119	AC105641 Rattus no
43	137.2	10.5	246196	2	AC105641	AE011088 Methanosa
44	95.8	7.3	10099	1	AE011088	AE011338 Leptospir
45	83	6.4	11938	1	AE011338	

ALIGNMENTS

RESULT 1
BD270234
LOCUS BD270234 1305 bp DNA linear PAT 17-JUL-2003
DEFINITION Human glycosylation enzymes.
ACCESSION BD270234
VERSION BD270234.1 GI:33080002
KEYWORDS JP 2002537796-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1305)
AUTHORS Coleman,T.A. and Betenbaugh,M.J.
TITLE Human glycosylation enzymes
JOURNAL Patent: JP 2002537796-A 1 12-NOV-2002;

HUMAN GENOME SCIENCES INC., JOHNS HOPKINS UNIVERSITY	
CS Homo sapiens (human)	
EV JP 2002537796-A/1	
PD 12-NOV-2002	
PF 01-MAR-2000 JP 2000602748	
PR 02-MAR-1999 US 60/122409	
PI TIMOTHY A COLEMAN, MICHAEL J SETENBAUGH	
PC C12N15/09, A61K31/715, A61K38/43, A61K48/00, A61P1/02, A61P3/00, PC A61P7/00, A61P17/02, A61P15/00, A61P13/12, A61P11/00, A61P17/02, PC A61P19/02, A61P21/04, A61P25/00, A61P25/16, A61P25/24, A61P25/28, A61P29/00, A61P31/18, A61P35/00, A61P37/00, A61P43/00, C07K16/40, C12N1/15, PC C12N1/19, C12N15/00, C12N9/00, G01N33/50, G01N33/53, PC C12N15/00, A61K37/48	
CC Human glycosylation enzymes	
FH Key Location/Qualifiers	
F7 CDS (1) . (1302)	
FEATURES	
source	1. 1305
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
ORIGIN	
Query Match	99.9%; Score 1303.4; DB 6; Length 1305;
Best Local Similarity	100.0%; Pred. No. 1.4e-301;
Matches 1305; Conservative	0; Mismatches C; Indels C; Gaps 0;
QY 1 ATGACCTCGTGGAGAGGGGGCCGACCTCGTCTCCAAACCCGCGGGGGCGACCGTCC 60	
DB 1 ATGACCTCGTGGAGAGGGGGCCGACCTCGTCTCCAAACCCGCGGGGGCGACCGTCC 60	
QY 61 CGGGCCCGCGGAGCTCAGGGGAACTCTCGCGGGCCGAGGGCGAGGTGTGGAG 120	
DB 61 CGGGCCCGCGGAGCTCAGGGGAACTCTCGCGGGCCGAGGGCGAGGTGTGGAG 120	
QY 121 AAGCCCGCGACCTGGAGCCCTAATCTGGCCCGGGGAGGAGCAAGGCAATCCCTCTG 180	
DB 121 AAGCCCGCGACCTGGAGCCCTAATCTGGCCCGGGGAGGAGCAAGGCAATCCCTCTG 180	
QY 181 AAGAACATTAAGCACTCGCGGGGGTCCCGTCTCATTTGGCTGGGTCTCGTGGCGCCCTG 240	
DB 181 AAGAACATTAAGCACTCGCGGGGGTCCCGTCTCATTTGGCTGGGTCTCGTGGCGCCCTG 240	
QY 241 GATTCAGGGGCTCCAGAGGTGATGGTTTCGACAGACCATGATGAATGAGAAATGTG 300	
DB 241 GATTCAGGGGCTCCAGAGGTGATGGTTTCGACAGACCATGATGAATGAGAAATGTG 300	
QY 301 GCCAAACANTTGGTGCAAGATTCATCGAAGAGTCTTGAAGTTTCAAAAGACAGCTCT 360	
DB 301 GCCAAACANTTGGTGCAAGATTCATCGAAGAGTCTTGAAGTTTCAAAAGACAGCTCT 360	
QY 361 ACCTCAGTAGTGCATCATAGAAATTTCTTAATTAATTAATTAATTAATTAATTAATTA 420	
DB 361 ACCTCAGTAGTGCATCATAGAAATTTCTTAATTAATTAATTAATTAATTAATTAATTA 420	
QY 421 AATATTCAGCTACTCTTCATGTTTACACCTACTGATCTTCAAAAAGTTTCAGAAATG 480	
DB 421 AATATTCAGCTACTCTTCATGTTTACACCTACTGATCTTCAAAAAGTTTCAGAAATG 480	
QY 481 ATTCCAGAAAGAGGATATGATTCGTTTCTCTGTTGTGAGACGCCATCAGTTTCGATGG 540	
DB 481 ATTCCAGAAAGAGGATATGATTCGTTTCTCTGTTGTGAGACGCCATCAGTTTCGATGG 540	
QY 541 AGTGAATTCAGAAAGGAGTTCGTGAAGTACCGAACCTCTGHTTTAATTCAGCTAAA 600	
DB 541 AGTGAATTCAGAAAGGAGTTCGTGAAGTACCGAACCTCTGHTTTAATTCAGCTAAA 600	
QY 601 CGGCTCTCGACAGAGCTGGGATGGAGAAATATATGAAATGGCTCATTTTATTTGCT 660	

DB 601 CGGCTCTCGACAGAGCTGGGATGGAGAAATATATGAAATGGCTCATTTTATTTGCT 660	
QY 661 AAAAGACATTTGATAGAGATGGGTACTTCAGAGGTGAAATGGCATACTACGAAATGC 720	
DB 661 AAAAGACATTTGATAGAGATGGGTACTTCAGAGGTGAAATGGCATACTACGAAATGC 720	
QY 721 GAGCTGACATAGTGTGGATATAGATGTGGATATTTGATTTGGCTATTTCAGAGCAAGA 780	
DB 721 GAGCTGACATAGTGTGGATATAGATGTGGATATTTGATTTGGCTATTTCAGAGCAAGA 780	
QY 781 GTATTAAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGAATAAACTTTTGGTTTC 840	
DB 781 GTATTAAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGAATAAACTTTTGGTTTC 840	
QY 841 AATATTGATGGATGCTCAGCAATGGCCCATTTATGTATCAGGAGACCAAAAGAAATA 900	
DB 841 AATATTGATGGATGCTCAGCAATGGCCCATTTATGTATCAGGAGACCAAAAGAAATA 900	
QY 901 ATATCTTATGATGTAAAGATGCTATTGGGATAGTTTATTTAAAGAAAGTGTATTGAG 960	
DB 901 ATATCTTATGATGTAAAGATGCTATTGGGATAGTTTATTTAAAGAAAGTGTATTGAG 960	
QY 961 GTGAGGCTAATCTCAGAAAGGCGCTGTCAAAGCAGAGCGTCTCTTTTAAACTGGAT 1020	
DB 961 GTGAGGCTAATCTCAGAAAGGCGCTGTCAAAGCAGAGCGTCTCTTTTAAACTGGAT 1020	
QY 1021 TGCAAAATGGAAGTCAGTGTATCAGACAGCTTGGATAGTTTATTTAAAGAAAGTGTATTGAG 1080	
DB 1021 TGCAAAATGGAAGTCAGTGTATCAGACAGCTTGGATAGTTTATTTAAAGAAAGTGTATTGAG 1080	
QY 1081 ATGGGCTCTGCTCGAAAGAGTGGCATATCTTGGAAATGAAGTGTCTGTATGAAGAGTGC 1140	
DB 1081 ATGGGCTCTGCTCGAAAGAGTGGCATATCTTGGAAATGAAGTGTCTGTATGAAGAGTGC 1140	
QY 1141 TTGAAGAGAGTGGGCTAAGTGGCGCTCTCTGCTATGCTGTCTTCTAGCCCGAGAGCT 1200	
DB 1141 TTGAAGAGAGTGGGCTAAGTGGCGCTCTCTGCTATGCTGTCTTCTAGCCCGAGAGCT 1200	
QY 1201 GTTGGATACATTTGCAATGTAATGGTGGCGGTGGCCATCCGAGAAATTTTCAGAGACAC 1260	
DB 1201 GTTGGATACATTTGCAATGTAATGGTGGCGGTGGCCATCCGAGAAATTTTCAGAGACAC 1260	
QY 1261 ATTTGCTCTACTAATGGAAGAGTTAACTAATTCATGCCAAAATAG 1305	
DB 1261 ATTTGCTCTACTAATGGAAGAGTTAACTAATTCATGCCAAAATAG 1305	
RESULT 2	
AF397212	1305 bp mRNA linear PRI 02-AUG-2002
LOCUS	Homo sapiens cytidine monophospho-sialic acid synthase mRNA,
DEFINITION	complete cds.
ACCESSION	AF397212
VERSION	AF397212.1 GI:22085789
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 1305)
JOURNAL	Lawrence, S.M., Huddleston, K.A., Tomiya, N., Nguyen, N., Lee, Y.C.,
REFERENCE	Vann, W.F., Coleman, T.A. and Betenbaugh, M.J.
AUTHORS	Cloning and Expression of Human Sialic Acid Pathway Genes to
	Generate CMP-Sialic Acids in Insect Cells
	Unpublished
	2 (bases 1 to 1305)
	Lawrence, S.M., Huddleston, K.A., Tomiya, N., Nguyen, N., Lee, Y.C.,
	Vann, W.F., Coleman, T.A. and Betenbaugh, M.J.
	Direct Submission
	Submitted (05-JUL-2001) Chemical Engineering, Johns Hopkins
	University, 3400 N. Charles St., Baltimore, MD 21218, USA
	Location/Qualifiers
FEATURES	

Thu May 13 11:53:01 2004

```

source
1. 1305
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. 1305
CDS
/note="Cmp-Sas; CMP-sialic acid synthase"
/codon_start=1
/product="cytidine monophospho-sialic acid synthase"
/protein_id="AA090580.1"
/db_xref="GI:22085790"
/translation="MDSVEKGAATSVNPRPRGRPRKLPQRNSGCGRGVKKPH
LAALILAGGSGKPLKNIKLHAGVPLIGLWLRALDSGAFQSVWSDHDEIENVAK
QFQAQVRRSSEKSDSTDIIEFLNYXNDIIGNIQTASCLHPTDQKVMK
1RESGYDSFVVRHQRWPEIOKGVREVDPLNINFAKPRQMDGELVENSFY
PAKRLHLMGYLQKGYMYEMRAHSHVDIDWDPIAEORLVRYGVFGKELKBIK
LLVNDIGLCTNGHLYSGDQKEIISYDVDAIGLISLLKSGIEVLISERACSKOTL
SSLKLDCKMEVSVDKLAVDWEKEMGLCKREVAIIGNEVSDDECLKRVLSGNPAD
ACSYAQKAVGYICKNGRGRGAIRFPAEHICLLXKVNNSQK"

ORIGIN
Query Match 98.1%; Score 1280.4; DB 9; Length 1305;
Best Local Similarity 99.8%; Pred. No. 4.5e-296;
Matches 1304; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 ATGGACTCGTGGAGAGGGGCGCCACCTCCGCTCCCAACCCCGGGGGCGACCGTCC 60
Db 1 ATGGACTCGTGGAGAGGGGCGCCACCTCCGCTCCCAACCCCGGGGGCGACCGTCC 60
QY 61 CGGGCCCGGCGCGAAGCTGAGCGCAACTCTCGCGGGCGGCGAGCGGAGTGGAG 120
Db 61 CGGGCCCGGCGCGAAGCTGAGCGCAACTCTCGCGGGCGGCGAGCGGAGTGGAG 120
QY 121 AAGCCCGCGCACTCGACGCCCTAATCTGGCGCGGGGAGGAGCAGAAAGGCATCC 180
Db 121 AAGCCCGCGCACTCGACGCCCTAATCTGGCGCGGGGAGGAGCAGAAAGGCATCC 180
QY 181 AAGAACATTAAGCACTGGCGGGGTCGGCTCATPCTGCTGGTCTCGGTGGCGGCTG 240
Db 181 AAGAACATTAAGCACTGGCGGGGTCGGCTCATPCTGCTGGTCTCGGTGGCGGCTG 240
QY 241 GATTCAGGGGCGCTTCAGAGTGTATGGTTCGACAGACATGATCAATGCGAATGG 300
Db 241 GATTCAGGGGCGCTTCAGAGTGTATGGTTCGACAGACATGATCAATGCGAATGG 300
QY 301 GCCAACAAATTTGGTCCACAAGTTTCATCGAAGAGTTCGAAGTTTCAAAAGACAGCT 360
Db 301 GCCAACAAATTTGGTCCACAAGTTTCATCGAAGAGTTCGAAGTTTCAAAAGACAGCT 360
QY 361 ACCTCACTAGATGCCATCATAGAAATTTCTTAATATATATATAGAGTGTAGGA 420
Db 361 ACCTCACTAGATGCCATCATAGAAATTTCTTAATATATATATAGAGTGTAGGA 420
QY 421 AATATTCAGACTACTCTTCATGCTTTACATCTCTACTGACTCTCAGAAAGTTG 480
Db 421 AATATTCAGACTACTCTTCATGCTTTACATCTCTACTGACTCTCAGAAAGTTG 480
QY 481 ATTTCAGAGAGAGATATGATCTCTGTTTCTGTTGTGAGAGCCCATGATTTCC 540
Db 481 ATTTCAGAGAGAGATATGATCTCTGTTTCTGTTGTGAGAGCCCATGATTTCC 540
QY 541 AGTGAATTCAGAAAGAGTTCGTGAAGTGCACCGAAGCTCTGAATTTAAATCC 600
Db 541 AGTGAATTCAGAAAGAGTTCGTGAAGTGCACCGAAGCTCTGAATTTAAATCC 600
QY 601 CGGCTCGTCGACAGACTGGATGAGATTTATGAAATGGCTCATTTTATTTGCT 660
Db 601 CGGCTCGTCGACAGACTGGATGAGATTTATGAAATGGCTCATTTTATTTGCT 660
QY 661 AAGAACATTTGATAGAGATGGGTTACTTCAGGGTGG-AAAATGGCATACTACGA 719
Db 661 AAGAACATTTGATAGAGATGGGTTACTTCAGGGTGG-AAAATGGCATACTACGA 720
QY 720 CGAGCTGGAAACATAGTSGGATATAGATGTTGATATTGATGGCCTATTGCGAG 779

```

```

Db 721 CGAGCT-GAACATAGTGTGGATATAGATGTGATATTGATTGGCCTATTGCGAGCA 779
QY 780 AGTATTAAGATATGCTATTTTGGCAAGAGAGAGCTTAAGGAAATAAATCTTTGTTG 839
Db 780 AGTATTAAGATATGCTATTTTGGCAAGAGAGAGCTTAAGGAAATAAATCTTTGTTG 839
QY 840 CAATATTGATGATGCTCACCACATGGCCACATTTATGATATCAGAGAGACCAAAAGAAAT 899
Db 840 CAATATTGATGATGCTCACCACATGGCCACATTTATGATATCAGAGAGACCAAAAGAAAT 899
QY 900 AATATCTTATGATGATAAAGATGCTATTGGGATAAGATTATTAAAGAAAAGTGGTATTC 959
Db 900 AATATCTTATGATGATAAAGATGCTATTGGGATAAGATTATTAAAGAAAAGTGGTATTC 959
QY 960 GGTGAGGCTTAATCTCAGAAAGGGCTGTTTAAAGCAGACGCTGCTCTTTTAAACTGGA 1019
Db 960 GGTGAGGCTTAATCTCAGAAAGGGCTGTTTAAAGCAGACGCTGCTCTTTTAAACTGGA 1019
QY 1020 TTGCAAAATGGAAGTCAGTGTATCAGACAGCTAGCTAGTGTAGATGAATGGAGAAAAGA 1079
Db 1020 TTGCAAAATGGAAGTCAGTGTATCAGACAGCTAGCTAGTGTAGATGAATGGAGAAAAGA 1079
QY 1080 AATGGCCCTGCTGCTGAAAGAGAGTGGCATATCTTGGAAATGAAGTCTCTGATCAAGAGTG 1139
Db 1080 AATGGCCCTGCTGCTGAAAGAGAGTGGCATATCTTGGAAATGAAGTCTCTGATCAAGAGTG 1139
QY 1140 CTTGAAGAGAGTGGGCTTAAGTGGGCTCCTGCTGATAGCTGTCTTACCCAGAGGC 1199
Db 1140 CTTGAAGAGAGTGGGCTTAAGTGGGCTCCTGCTGATAGCTGTCTTACCCAGAGGC 1199
QY 1200 TGTGATGATCATTTGCAAAATTAAGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1259
Db 1200 TGTGATGATCATTTGCAAAATTAAGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1259
QY 1260 CATTTGCTCTACTAATGGAAGAAAGTTAATAATTCATGCCAAAATAG 1305
Db 1260 CATTTGCTCTACTAATGGAAGAAAGTTAATAATTCATGCCAAAATAG 1305

```

```

RESULT 3
AF271388 1376 bp mRNA linear PRI 14-JUN-2000
LOCUS Homo sapiens CMP-N-acetylneuraminic acid synthase mRNA, complete
DEFINITION cds.
ACCESSION AF271388
VERSION AF271388.1 GI:8515842
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1376)
AUTHORS Bouquin,T. and Mundy,J.
TITLE Human mRNA for CMP-N-acetylneuraminic acid synthase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1376)
AUTHORS Bouquin,T. and Mundy,J.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2000) Department of Plant Physiology, Molecular
Biology Institute, Oester Farimagsgade 2A, Copenhagen 2353 K,
Denmark
FEATURES
source Location/Qualifiers
1. 1376
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
19. 1323
/EC_number="2.7.7.43"
/codon_start=1
/product="CMP-N-acetylneuraminic acid synthase"
/protein_id="AAF76203.1"
CDS

```

/db_xref="GI:8515843"
/translation="MDSVEKGAATSVSNPRGRPRPPKLRNSRGGQGRVKEPPH
LAALLARGSGKIPKNIKLAGVPLIGWLVNEDIVGNIOATSPCLPTDILKQVAK
QFGAQRSSRSVSDSTSJDAIIEFLVNEVDEPLNPAKPRRODMDELXENGSEFY
IKREGYDSVSVRRHQRWSEKQGVRETEPLNPAKPRRODMDELXENGSEFY
FAKHLIEMGYLGGKMYDREAHSDVIDIDWPIAEORVLYAGVYGEKLEIK
LIVNIDGLTNGHIVGSDQKEIISYVDKDAIGLSLLKSGVRLISERACSKQTL
SSLKDCMEVSVSKLAVVDEWREKMLGWEVAYLGNESVDEBCKRKLVLGSGAPAD
ACSTAQKAVGYICKNGRGRGAREFAEHICLMEKVNNSQK"

ORIGIN

Query Match 97.5%; Score 1272.4; DB 3; Length 1376;
Best Local Similarity 99.2%; Pred. No. 3.9e-294;
Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;
QY 1 ATGGACTCGGTGAGAGGGGGCCGACCTCGTCTCCAAACCGCGGGGGGACCGTCC 60
DB 19 ATGGACTCGGTGAGAGGGGGCCGACCTCGTCTCCAAACCGCGGGGGGACCGTCC 78
QY 61 CGGGGGCGGGCCGCGAAGCTGCAGCGCAACTCTCGCGGGGGCCAGGGCCGAGGTGGAG 120
DB 73 CGGGGGCGGGCCGCGAAGCTGCAGCGCAACTCTCGCGGGGGCCAGGGCCGAGGTGGAG 138
QY 121 AAGCCCCCGCACTGGCAGCCCTAATCTGGCCCGGGGGAGGAGCAAGGCATCCCGCTG 180
DB 139 AAGCCCCCGCACTGGCAGCCCTAATCTGGCCCGGGGGAGGAGCAAGGCATCCCGCTG 198
QY 181 AAGAACATTAAGCACCTGGCGGGGGTCCCGCTCATTTGGCTGGGTCTGCGCGCCCTG 240
DB 199 AAGAACATTAAGCACCTGGCGGGGGTCCCGCTCATTTGGCTGGGTCTGCGCGCCCTG 258
QY 241 GATTCAGGGGCCCTCCAGAGTGTATGGTTTCGACGACCATCATGTAATTTGAGATGTG 300
DB 259 GATTCAGGGGCCCTCCAGAGTGTATGGTTTCGACGACCATCATGTAATTTGAGATSTG 318
QY 301 SCCAACAATTTGGTGACAGTTCATCGAGAGTTCGAGTTTCAAAGAGCAGCTCT 360
DB 319 SCCAACAATTTGGTGACAGTTCATCGAGAGTTCGAGTTTCAAAGAGCAGCTCT 378
QY 361 ACCTCAGTAGATGCATCATAGATTTCTTAATTAATTAATGAGGKTCATGTAGGA 420
DB 379 ACCTCAGTAGATGCATCATAGATTTCTTAATTAATTAATGAGGTTGATGTAGGA 438
QY 421 AATATTCAGCTACTTCTYCATGTTTACATCTTCTTCAAAAGTTCGAGAAATG 480
DB 439 AATATTCAGCTACTTCTTCCATGTTTACATCTTCTTCAAAAGTTCGAGAAATG 498
QY 481 ATTTCGAGAGAGATATGATTCCTGTTCTGTTGAGACCCCATCAGTTTCGATGG 540
DB 499 ATTTCGAGAGAGATATGATTCCTGTTCTGTTGAGACCCCATCAGTTTCGATGG 558
QY 541 AATGAAATTCAGAAAGAGTTCGTAAGTGCAGGAGTCCGAACTCTGAAATTCAGCTTAA 600
DB 559 AATGAAATTCAGAAAGAGTTCGTAAGTGCAGGAGTCCGAACTCTGAAATTCAGCTTAA 618
QY 601 CGGCTCTGTCAGACAGCTGGATGAGAAATTAATGAAATGCTTCAATTTATTTGCT 660
DB 619 CGGCTCTGTCAGACAGCTGGATGAGAAATTAATGAAATGCTTCAATTTATTTGCT 678
QY 661 AAAAGACATTTGATAGAGATGGTTTACTTCAGGGTGG-AAAATGGCATCTACGAAATG 719
DB 679 AAAAGACATTTGATAGAGATGGTTTACTTCAGGGTGG-AAAATGGCATCTACGAAATG 738
QY 720 CGAGCTGGAACATAGTGTGGATATAGATGTGGATATGATTTGGGCTTATTCAGAGCAAG 779
DB 739 CGAGCT-GAACATAGTGTGGATATAGATGTGGATATGATTTGGGCTTATTCAGAGCAAG 797
QY 780 AGTAAATAGATGCTATTTTGGCAAGAGAGCTTAAGGAAATAAAGCTTTTGGTTG 839
DB 798 AGTAAATAGATGCTATTTTGGCAAGAGAGCTTAAGGAAATAAAGCTTTTGGTTG 857
QY 840 CAATATTCATGCTCTCCACCAATGGCCACATTTATGATCAGGAGACCAAAAGAAAT 899

DB 858 CAATATTCATGCTCTCCACCAATGGCCACATTTATGATCAGGAGACCAAAAGAAAT 917
QY 900 AATATCTTATGATGTAAGAGATGCTATTTGGGTAAGATTTATTAAGAAAAGTGGTATGA 959
DB 918 AATATCTTATGATGTAAGAGATGCTATTTGGGTAAGATTTATTAAGAAAAGTGGTATGA 977
QY 960 GGTGAGGCTAATCTCAGAAAGGGCCCTGTCTCAAGCAGACGCTGTCTTTTAAATCGGA 1019
DB 978 GGTGAGGCTAATCTCAGAAAGGGCCCTGTCTCAAGCAGACGCTGTCTTTTAAATCGGA 1037
QY 1020 TTGCAAAATGGAAGTCACTGTATCAGACAGCTAGCTAGTGTAGATGAGTGGAGAAAGA 1079
DB 1038 TTGCAAAATGGAAGTCACTGTATCAGACAGCTAGCTAGTGTAGATGAGTGGAGAAAGA 1097
QY 1080 AATGGCCCTGTCTGAAAGAGAGTGGCATATCTTGAAATGAAAGTGTCTGATGAGAGTG 1139
DB 1098 AATGGCCCTGTCTGAAAGAGAGTGGCATATCTTGAAATGAAAGTGTCTGATGAGAGTG 1157
QY 1140 CTTGAAAGAGAGTGGCCCTAAGTGGCCCTCTGCTGATGCTGTCTTCTACGCCAGAGGC 1199
DB 1158 CTTGAAAGAGAGTGGCCCTAAGTGGCCCTCTGCTGATGCTGTCTTCTACGCCAGAGGC 1217
QY 1200 TGTGATACATTTGCAAAATGATGCTGGCCCTGCTGCTGATGCTGTCTTCTACGCCAGAGGC 1259
DB 1218 TGTGATACATTTGCAAAATGATGCTGGCCCTGCTGCTGATGCTGTCTTCTACGCCAGAGGC 1277
QY 1260 CATTTGCTACTAATGSAAGTAAATATTCATGCCAAAATAG 1305
DB 1278 CATTTGCTACTAATGSAAGTAAATATTCATGCCAAAATAG 1323

RESULT 4
HSM804286 1766 bp mRNA linear PRI 10-JUL-2002
LOCUS Homo sapiens mRNA; cDNA DKFp666I142 (from clone DKFp666I142).
DEFINITION AL832975
ACCESSION AL832975.1 GI:21733563
VERSION
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1766)
AUTHORS Ansoxge, W., Wirkner, U., Mewes, H.W., Weil, B. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1. D-85764 Neuberberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFp666I142) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.

FEATURES
1. 1766
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFp666I142"
/db_xref="taxon:9606"
/clone="DKFp666I142"
/tissue type="stomach"
/clone_lib="666 (synonym: hsto2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev stage="adult"
1714..1719
1731
polyA signal
polyA_site
ORIGIN
Query Match 97.5%; Score 1272.4; DB 9; Length 1766;
Best Local Similarity 99.2%; Pred. No. 3.9e-294;

Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;

QY 1 ATGGACTCGGTGGAGAGGGGGCCGCCACCTCCCGTCTCCAAACCCCGGGGGGACCGTCC 60
 DB ATGGACTCGGTGGAGAGGGGGCCGCCACCTCCCGTCTCCAAACCCCGGGGGGACCGTCC 138
 QY 61 CGGGCCCGCCCGCGAAGCTGCGAGCGCAACTCTCGCGGGCGGCAAGGCGAGGTGGAG 120
 DB CGGGCCCGCCCGCGAAGCTGCGAGCGCAACTCTCGCGGGCGGCAAGGCGAGGTGGAG 198
 QY 121 AAGCCCCCGACCTCGAGACCTTAATCTGCGCGGGGAGGAGCAAGGCAATCCCGCTG 180
 DB AAGCCCCCGACCTCGAGACCTTAATCTGCGCGGGGAGGAGCAAGGCAATCCCGCTG 258
 QY 181 AAGAACATTAAGCACTGCGCGGGGTCCCGCTCAITGGCTGGGTCTCTGGTGGTCTG 240
 DB AAGAACATTAAGCACTGCGCGGGGTCCCGCTCAITGGCTGGGTCTCTGGTGGTCTG 318
 QY 241 GATTCAGGGGCTTCCAGAGTGTATGGGTTCGACAGACCATGATGAATTCAGAAATG 300
 DB GATTCAGGGGCTTCCAGAGTGTATGGGTTCGACAGACCATGATGAATTCAGAAATG 378
 QY 301 GCGAAACAAATTTGGTGCACAGTTCATCGAAGATTTCTGAAGTTTCAAAAGACAGCTCT 360
 DB GCGAAACAAATTTGGTGCACAGTTCATCGAAGATTTCTGAAGTTTCAAAAGACAGCTCT 438
 QY 361 NCTCCTAGATGCCATCATAGATTTCTTAATTAATTAATGAAGGKTCAGATTTAGGA 420
 DB NCTCCTAGATGCCATCATAGATTTCTTAATTAATTAATGAAGGKTCAGATTTAGGA 498
 QY 421 AATATTCAAGCTACTTCTCCTCAITGTTTACATCTTACTGATCTTCAAAAGTTG 480
 DB AATATTCAAGCTACTTCTCCTCAITGTTTACATCTTACTGATCTTCAAAAGTTG 558
 QY 481 ATTGAGAAAGAGGATATGATTTCTGTTTCTGTTGTCAGAGCCATCAGTTTCATG 540
 DB ATTGAGAAAGAGGATATGATTTCTGTTTCTGTTGTCAGAGCCATCAGTTTCATG 618
 QY 541 ACTGAAATTCAGAAAGAGTCTGTGAAGTGACCGAACCTCTGTAATTTAAATCCAGCTAA 600
 DB ACTGAAATTCAGAAAGAGTCTGTGAAGTGACCGAACCTCTGTAATTTAAATCCAGCTAA 678
 QY 601 CGGCTCTGTCGACAGACTGGATGGAGATTAATATGAATGGCTCAATTTATTTGCT 660
 DB CGGCTCTGTCGACAGACTGGATGGAGATTAATATGAATGGCTCAATTTATTTGCT 738
 QY 661 AAAAGCAATTTGATAGAGATGGTTACTTTCAGAGTGG-AAAATGTCATACAGAAATG 719
 DB AAAAGCAATTTGATAGAGATGGTTACTTTCAGAGTGG-AAAATGTCATACAGAAATG 798
 QY 720 CGAGCTGGAACATAGTGTGGATATAGATGTGGATATTTGATTTGCTTATTCAGAGCAAG 779
 DB CGAGCT- GAAATAGTGTGGATATAGTGTGGATATTTGATTTGCTTATTCAGAGCAAG 857
 QY 780 AGTATTAAGATATGCTATTTTGGGAAAGAGCTTACAGAAATTAACCTTTGTTG 839
 DB AGTATTAAGATATGCTATTTTGGGAAAGAGCTTACAGAAATTAACCTTTGTTG 917
 QY 840 CAATATGATGATGCTTCCCAATGCGCACATTTATGATATCAGGAGCCAAAGAAAT 899
 DB CAATATGATGATGCTTCCCAATGCGCACATTTATGATATCAGGAGCCAAAGAAAT 977
 QY 900 AATCATCTTATGATTAAGATGCTATTTGGATAAGTTTATTAAGAAAGAGTGGTATGA 959
 DB AATCATCTTATGATTAAGATGCTATTTGGATAAGTTTATTAAGAAAGAGTGGTATGA 1037
 QY 960 GGTGAGCTAATCTCAGAAAGGGCTGTTTCAAGAGCAGACGCTGCTCTTTTAAACCTGGA 1019
 DB GGTGAGCTAATCTCAGAAAGGGCTGTTTCAAGAGCAGACGCTGCTCTTTTAAACCTGGA 1097
 QY 1020 TTGCAAAATGGAAGTCAAGTATCAGCAAGCTAGCAGATTTGATGATCAATGGAAGAA 1079
 DB TTGCAAAATGGAAGTCAAGTATCAGCAAGCTAGCAGATTTGATGATCAATGGAAGAA 1157

QY 1080 AATGGCCCTGCTGCTGSHAAAGAGTGGCATATCTTGAATGAAGTCTCTGATGAAGATG 1139
 DB 1158 AATGGCCCTGCTGCTGSHAAAGAGTGGCATATCTTGAATGAAGTCTCTGATGAAGATG 1217
 QY 1140 CTGTAAGAGAGTGGGCTTAAGTGGGCTCTCTGCTGATGCTGTTCTACGCCAGAGGC 1199
 DB 1218 CTGTAAGAGAGTGGGCTTAAGTGGGCTCTCTGCTGATGCTGTTCTACTGCCAGAGGC 1277
 QY 1200 TCTTGGATACATTTGCAATGTAATGGTGGCGCTGGTGGTCCCTCCGAGATTTGCAGACA 1259
 DB 1278 TCTTGGATACATTTGCAATGTAATGGTGGCGCTGGTGGTCCCTCCGAGATTTGCAGACA 1337
 QY 1260 CATTTGCTTACTAATGGAAGATTTAATTAATTCATGCCAAATAAG 1305
 DB 1338 CATTTGCTTACTAATGGAAGATTTAATTAATTCATGCCAAATAAG 1383

RESULT 5
 AR338869
 LOCUS
 REFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 1 (bases 1 to 1770)
 AUTHORS
 Tang, Y. T., Zhou, P., and Drmanac, R. T.
 TITLE
 Nucleic acids and polypeptides
 JOURNAL
 Patent: US 659662-A 360 27-MAY-2003;
 FEATURES
 1..1770
 Location/Qualifiers
 /organism="unknown"
 /mol_type="genomic DNA"

Query Match 97.5%; Score 1272.4; DB 6; Length 1770;
 Best Local Similarity 99.2%; Pred. No. 4e-294;
 Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;

QY 1 ATGGACTCGGTGGAGAGGGGGCCGCCACCTCCCGTCTCCAAACCCCGGGGGGACCGTCC 60
 DB 105 ATGGACTCGGTGGAGAGGGGGCCGCCACCTCCCGTCTCCAAACCCCGGGGGGACCGTCC 164
 QY 61 CGGGCCCGCCCGCGAAGCTGCGAGCGCAACTCTCGGGCGGCGGAGGCTGGAG 120
 DB 165 CGGGCCCGCCCGCGAAGCTGCGAGCGCAACTCTCGGGCGGCGGAGGCTGGAG 224
 QY 121 AAGCCCCCGACCTGCGAGACCTTAATTTCTGCGCGGGGAGGAGCAAGGCAATCCCGCTG 180
 DB 225 AAGCCCCCGACCTGCGAGACCTTAATTTCTGCGCGGGGAGGAGCAAGGCAATCCCGCTG 284
 QY 181 AAGAAATTAAGCACTGCGCGGGGTCCCGCTCAITGGCTGGGTCTCTGGTGGTCTG 240
 DB 285 AAGAAATTAAGCACTGCGCGGGGTCCCGCTCAITGGCTGGGTCTCTGGTGGTCTG 344
 QY 241 GATTCAGGGGCTTCCAGAGTGTATGGGTTCGACAGACCATGATGAATTCAGAAATG 300
 DB 345 GATTCAGGGGCTTCCAGAGTGTATGGGTTCGACAGACCATGATGAATTCAGAAATG 404
 QY 301 GCGAAACAAATTTGGTGCACAGTTCATCGAAGATTTCTGAAGTTTCAAAAGACAGCTCT 360
 DB 405 GCGAAACAAATTTGGTGCACAGTTCATCGAAGATTTCTGAAGTTTCAAAAGACAGCTCT 464
 QY 361 ACCTCATAGATGCCATCAGAGATTTCTTAATTAATTAATGAAGTGAATTTAGGA 420
 DB 465 ACCTCATAGATGCCATCAGAGATTTCTTAATTAATTAATGAAGTGAATTTAGGA 524
 QY 421 AATATTCAAGCTACTTCTCCTCAITGTTTACATCTTACTGATCTTCAAAAGTTG 480
 DB 525 AATATTCAAGCTACTTCTCCTCAITGTTTACATCTTACTGATCTTCAAAAGTTG 584

JOURNAL Patent: EP 1074617-A 17324 07-FEB-2001; Research Association for Biotechnology (JP)

FEATURES

```

source
1. .1345
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
41. .1345
CDS
/feature="unassigned protein product"

```

ORIGIN

Query Match	97.4%	Score 1270.8;	DB 6;	Length 1694;
Best Local Similarity	99.1%	Pred. No. 9.5e-294;		
Matches 1294:	Conservative	4;	Mismatches 6;	Indels 2;
				Gaps 2;

QY	1	ATGGACTCGTGTGAGAAAGGGGGCGGCACTTCGTTCTTCAACCCGCGGGGGGACCGTCC	60
Db	41	ATGGACTCGTGTGAGAAAGGGGGCGGCACTTCGTTCTTCAACCCGCGGGGGGACCGTCC	100
QY	61	CGGGGGCGGGCGGCGGAACTGACAGCGCMACTTCGCGGGCGGCAGGGCGAGGTGTGGAG	120
Db	101	CGGGGGCGGGCGGCGGAACTGCGAGCTGCGAGCACTTCGCGGGCGGCAGGGCGAGGTGTGGAG	160
QY	121	AAGCCCCGCGACCTTGGCAGCGCTTAATTCTGGCCCGGGGAGGCGAGCAAGGCATCCCCCTG	180
Db	161	AAGCCCCGCGACCTTGGCAGCGCTTAATTCTGGCCCGGGGAGGCGAGCAAGGCATCCCCCTG	220
QY	181	AAGAACTTAAGCACCTTGGGGGGGTCGCCCTCATTTGGCTGGGTTCCTCGTGGCGGCTGTG	240
Db	221	AAGAACTTAAGCACCTTGGGGGGGTCGCCCTCATTTGGCTGGGTTCCTCGTGGCGGCTGTG	280
QY	241	GATTCAGGGGCTTCCAGAGTGTATGGTTTCGNCAGACCATGATGAATTTGAGAATGTG	300
Db	281	GATTCAGGGGCTTCCAGAGTGTATGGTTTCGACAGCCATGATGAATTTGAGAATGTG	340
QY	301	GCACAACAATTTGGTGCACAAAGTTCATCGAAGAAGTTCTGAAGTTTCAAAAAGACAGCTCT	360
Db	341	GCCAAACAATTTGGTGCACAAAGTTCATCGAAGAAGTTCTGAATTTCAAAAAGACAGCTCT	400
QY	361	ACCTCACTAGATGCCATCATAAGATTTCTTAATTTATVACAATGAGGKTGCATTTGTAGGA	420
Db	401	ACCTCACTAGATGCCATCATAAGATTTCTTAATTTATCATAATGAGGTTTACATTTGTAGGA	460
QY	421	AATATCTCAAGCTACTTCTTCYCAATGTTTACATCCTTACTGATCTTCAAAAAGTTCGACAAATG	480
Db	461	AATATCTCAAGCTACTTCTTCCATGTTTATCATCCTTACTGATCTTCAAAAAGTTCGACAAATG	520
QY	481	ATTCGAGNAGAAGGATATGATTTCTGTTGTTGAGACGTCATCAGTTTCGATGG	540
Db	521	ATTCGAGNAGAAGGATATGATTTCTGTTGTTGAGACGTCATCAGTTTCGATGG	580
QY	541	AGTCAAAATTCAGAAAGGAGTTTCGTGAAGTGACCGAACTCTCAATTTAAATCCAGCTAAA	600
Db	581	AGTCAAAATTCAGAAAGGAGTTTCGTGAAGTGACCGAACTCTCTGAATTTAAATCCAGCTAAA	640
QY	601	CGGCGCTCGTCACAAGACTGGGATGGAGAAATTATATGAAATTGGCTCAATTTATTTTCCT	660
Db	641	CGGCGCTCGTCACAAGACTGGGATGGAGAAATTATGAAATTGGCTCAATTTATTTTCCT	700
QY	661	AAAAGACATTTGATAGAGATGGGTTACTTGCAGGGTGG-AAAAATGSCATACTACGAAATG	719
Db	701	AAAAGACATTTGATAGAGATGGGTTACTTGCAGGGTGGTAC-TTCAGGGTGGAAATAATGCAATCATAGAAATG	760

481	ATTGAGAAAGAGGATCTGKTTCTCTCTGTGTGAGAGCCCATCAGTTCGATGG	540	QY
585	ATTGAGAAAGAGGATATGATTTCTGTGTGTGAGAGCCCATCAGTTCGATGG	644	Db
541	AGTGAATTCAGAAAGGAGTTTCGTGACGTACCGAACCTCTGAATTTAAATCCAGCTAAA	600	QY
645	AGTGAATTCAGAAAGGAGTTTCGTGACGTACCGAACCTCTGAATTTAAATCCAGCTAAA	704	Db
601	CGGCCCTCTGCAACAAGACTGGGATGGAGAAATTATATGAAATTCGCTCATTTTATTTGCT	660	QY
705	CGGCCCTCTGCAACAAGACTGGGATGGAGAAATTATATGAAATTCGCTCATTTTATTTGCT	764	Db
661	AAAGACATTTGATAGAGATGGTTACTTCGACGGGTGG-AAAATGGCATCTACGAAATTS	719	QY
765	AAAGACATTTGATAGAGATGGTTACTTCGACGGGTGGAAATATGGCATCTACGAAATG	824	Db
720	CGAGCTGGAAACATAGTGTGGATATAGATGTGGATATTCATTTGGCCCTATTTCAGAGACAAAG	779	QY
825	CGAGCT-GAACATAGTGTGGATATAGATGTGGATATTCATTTGGCCCTATTTCAGAGACAAAG	883	Db
780	AGTATTAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGAAATATAAATCTTTGGTTTG	839	QY
884	AGTATTAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGAAATATAAATCTTTGGTTTG	943	Db
840	CAATATTGTGATGCTCACCATGGCCACATTTATCTGATCAGGAGACCAAAAAGAAAT	899	QY
944	CAATATTGTGATGCTCACCATGGCCACATTTATGTTATCAGGAGACCAAAAAGAAAT	1003	Db
900	AAATATCTTATGATGTAAAAGATGCTATTGGGATAAGTTTATTAAAGAAAGTGGTATTGA	959	QY
1004	AAATATCTTATGATGTAAAAGATGCTATTGGGATAAGTTTATTAAAGAAAGTGGTATTGA	1063	Db
960	GGTAGGCTTAATCTAGAAAGGGCCCTGTTCAAGCHAGAGCTGTCTTCTTTAAACTGGA	1019	QY
1064	GGTAGGCTTAATCTAGAAAGGGCCCTGTTCAAGCHAGAGAGCTGTCTTCTTTAAACTGGA	1123	Db
1020	TTTCRAAATGGAGTTCAGTCTCTCAGACAGCTAGCAGTTGTAGATGAATCGAGAAAAGA	1079	QY
1124	TTTCRAAATGGAGTTCAGTCTCTCAGACAGCTAGCAGTTGTAGATGAATCGAGAAAAGA	1183	Db
1080	AAATGGGCCCTGTGCTGGAAAGAAAGTGGCATCTTTGGAAATCAAGTCTCTGATGAAGAGTGG	1139	QY
1184	AAATGGGCCCTGTGCTGGAAAGAAAGTGGCATCTTTGGAAATCAAGTCTCTGATGAAGAGTGG	1243	Db
1140	CTTGAAGAGAGTGGGCCCTAAGTGGCGCTCTCTGCTGATGCTCTTCTCTACGCCCAGAGGC	1199	QY
1244	CTTGAAGAGAGTGGGCCCTAAGTGGCGCTCTCTGCTGATGCTCTTCTCTACTGCCCAGAGGC	1303	Db
1200	TGTTGGATACATTTGCAAAATGTAATGTGTGCGGTGGTGCCATCCGAGAAATTTTCAGAGCA	1259	QY
1304	TGTTGGATACATTTGCAAAATGTAATGTGTGCGGTGGTGCCATCCGAGAAATTTTCAGAGCA	1363	Db
1260	CATTGCGCTACTAATGAAAAAGTTAATAATATCATGCCAAAAATAG	1305	QY
1364	CATTGCGCTACTAATGAAAAAGTTAATAATATCATGCCAAAAATAG	1409	Db

RESULT 6					
AX8824-9					
LOCUS	AX882419	1694 bp.	DNA	linear	PAT 17-DEC-2003
DEFINITION	Sequence 17324 from Patent EP104617.				
ACCESSION	AX882419				
VERSION	AX882419.1	GI:40037278			
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
REFERENCE	1 Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T. Primers for synthesising full-length cDNA and their use				
AUTHORS					
TITLE					


```

340 AATATCTTATGATGTAAGAGATGCTATTGGATAGATTATTAAAGAAAGTGGTATTGA 999
960 GGTGAGGCTAATCTCAGAAAGGGCGCTGTTCAAAGCAGACGCTGCTCTTTAAACTGGA 1019
1000 GGTGAGGCTAATCTCAGAAAGGGCGCTGTTCAAAGCAGACGCTGCTCTTTAAACTGGA 1059
1020 TTGCAAAATGGAAGTCAAGTGTATCAGACAGACTAGCAGTGTAGATGAATGAGAAAGA 1079
1060 TTGCAAAATGGAAGTCAAGTGTATCAGACAGACTAGCAGTGTAGATGAATGAGAAAGA 1119
1080 AATGGGCGCTGTCGTGGAAAGAGTGGCATATCTTGSAAATGAAGTGTCTGATGAAGAGTG 1139
1120 AATGGGCGCTGTCGTGGAAAGAGTGGCATATCTTGSAAATGAAGTGTCTGATGAAGAGTG 1179
1140 CTTGAAGAGAGTGGGCTTAAGTGGCGCTCCTGCTGATGCTGTTCTAGCCCGCAGAGGC 1199
1180 CTTGAAGAGAGTGGGCTTAAGTGGCGCTCCTGCTGATGCTGTTCTAGCCCGCAGAGGC 1239
1200 TGTGGATACATTTGCAAAATGAATGTGGCGTGTGCCATCCAGAGATTTGCAGAGCA 1259
1240 TGTGGATACATTTGCAAAATGAATGTGGCGTGTGCCATCCAGAGATTTGCAGAGCA 1299
1260 CATTGGCTTACTAATGGAAAAAGTTAAATAATTCATGCCAAAAATAG 1305
1300 CATTGGCTTACTAATGGAAAAAGTTAAATAATTCATGCCAAAAATAG 1345

AK022927 1694 bp mRNA linear PRI 01-AUG-2002
Homo sapiens cDNA FLJ12865 fis. clone NT2RP2003643, moderately
similar to Mus musculus mRNA for CMP-N-acetylneuraminic acid
synthetase.
AK022927
AK022927.1 GI:10434598
oligo capping; fis (full insert sequence).
Homo sapiens (human)
Homo sapiens
Homo sapiens
Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 Isogai,T., Ota,T., Hayaashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,
Arita,M., MabeKura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Oshima,A.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 1694)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- and 3'-end one pass sequencing and clone selection
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
Location/Qualifiers
1. .1694
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP2003643"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_id="NT2RP2"
/note="Cloning vector: pME18SFL3-mRNA from NT2 neuronal
(RA)

```

```
/organism="Homo sapiens"  
/mol_type="genomic DNA"  
/db_xref="taxon:9606"
```

ORIGIN

Query Match	97.4%;	Score 1270.8;	DB 6;	Length 1772;
Best Local Similarity	99.1%;	Pred. No. 9.6e-294;		
Matches 1294;	Conservative	4;	Mismatches 6;	Indels 2; Gaps 2;
QY	1	ATGGACTCGGTGAGAAAGGGCGCGCACACTCGTCTCCAAACCCGCGGGGGGACCGTCC	60	
DB	101	ATGGACTCGGTGAGAAAGGGCGCGCACACTCGTCTCCAAACCCGCGGGGGGACCGTCC	160	
QY	61	CGGGGCGGCGCGGAAGCTGCAGGCGAACTCTCGCGGGCGCGCAGGCCCGAGTGTGGAG	120	
DB	161	CGGGGCGGCGCGGAAGCTGCAGGCGAACTCTCGCGGGCGCGCAGGCCCGAGTGTGGAG	220	
QY	121	AAGCCCCGACCTGCGAGCCCTAATCTTGGCCGGGGAGGCGCAAGGCAATCCCCCTG	180	
DB	221	AAGCCCCGACCTGCGAGCCCTAATCTTGGCCGGGGAGGCGCAAGGCAATCCCCCTG	280	
QY	181	AAGAACATTAAGCACCTGGCGGGGCTCCGCTCAATTGGCTGGGTCTCGTGGCGCCCTG	240	
DB	281	AAGAACATTAAGCACCTGGCGGGGCTCCGCTCAATTGGCTGGGTCTCGTGGCGCCCTG	340	
QY	241	GATTTCAGGGGCTTCAGAGTGTATGGTTCAGACAGCCTCATGATGAATGAGANTGTG	300	
DB	341	GATTTCAGGGGCTTCAGAGTGTATGGTTCAGACAGCCTCATGATGAATGAGANTGTG	400	
QY	301	GCCAAACAAATTCGTSCACAAAGTTCATCGAAGAAGTTCGAAGTTTCAAAGAAGCAGCTCT	360	
DB	401	GCCAAACAAATTCGTSCACAAAGTTCATCGAAGAAGTTCGAAGTTTCAAAGAAGCAGCTCT	460	
QY	361	ACCTCCTAGATGCCATPACATAGAAATTCCTTAAATATATATGAGGKTGACATTTGAGGA	420	
DB	461	ACCTCCTAGATGCCATPACATAGAAATTCCTTAAATATATGAGGKTGACATTTGAGGA	520	
QY	421	AATATTCAGCTACTCTTCATGTTTACATCTTACTGATCTTCAAAGTTGCGAAGATG	480	
DB	521	AATATTCAGCTACTCTTCATGTTTACATCTTACTGATCTTCAAAGTTGCGAAGATG	580	
QY	481	ATTCGAGAAGAAGATATGATTCCTGKTTTCTCTCTTGTGACAGCCCATCAGTTTCGATGG	540	
DB	581	ATTCGAGAAGAAGATATGATTCCTGKTTTCTCTCTTGTGACAGCCCATCAGTTTCGATGG	640	
QY	541	AGTGAATATCAGAAAGAGTTCGTGAGTGTGACCGAACTCTGGAATTTAAATCCAGCTAAA	600	
DB	641	AGTGAATATCAGAAAGAGTTCGTGAGTGTGACCGAACTCTGGAATTTAAATCCAGCTAAA	700	
QY	601	CGGCTCTGTCGACAAAGCTGGAGTGAAGATATATATGAAGATGGCTCATTTATTTTCT	660	
DB	701	CGGCTCTGTCGACAAAGCTGGAGTGAAGATATATATGAAGATGGCTCATTTATTTTCT	760	
QY	661	AAAGACATTTGATAGAGATGGGTACTTCAGGGTGG-AAAATGGCATCTACGAAATG	719	
DB	761	AAAGACATTTGATAGAGATGGGTACTTCAGGGTGG-AAAATGGCATCTACGAAATG	820	
QY	720	CGAGCTGAAACATAGTGTGGATATAGATGTGATATTCATTTGGCCCTATTTCAGAGCAAG	779	
DB	821	CGAGCT-GAACATAGTGTGGATATAGATGTGATATTCATTTGGCCCTATTTCAGAGCAAG	879	
QY	780	AGTATTAAGATATGCTATTTTGGCAAGAGAGCTTAAGGAAATATAACTTTTGGTTTG	839	
DB	880	AGTATTAAGATATGCTATTTTGGCAAGAGAGCTTAAGGAAATATAACTTTTGGTTTG	939	
QY	840	CAATATTCATGATGTCTCCACAAATGGGCCACATTTATGATCAGAGACCAAAAAGAAAT	899	
DB	940	CAATATTCATGATGTCTCCACAAATGGGCCACATTTATGATCAGAGACCAAAAAGAAAT	999	
QY	900	AATATCTTATGATGTAAAGATGCTATTTGGGATAGCTTATTATTAAGAAAGTGTATTGCA	959	
DB	1000	AATATCTTATGATGTAAAGATGCTATTTGGGATAGCTTATTATTAAGAAAGTGTATTGCA	105	

820	DB	AGTATTAAGATATGCTATTTTGGCAAGAGAGCTTAAGGAAATAAACACTTTGGTTG	879
840	QY	CAATATTGATGATGCTCACCAAATGGCCACATTTATGTATCAGGAGACCCAAAAGAAAT	899
880	DB	CAATATTGATGATGCTCACCAAATGGCCACATTTATGTATCAGGAGACCCAAAAGAAAT	939
900	QY	AATATCTTATGATGATAAAGATGCTATTGGGTAAGTATTATTAAGAAAAGTGGTATTGA	959
940	DB	AATATCTTATGATGATAAAGATGCTATTGGGTAAGTATTATTAAGAAAAGTGGTATTGA	999
960	QY	GGTAGGCTTAATCTCAGAAAGGGCCGTTCAAAGCAGACGCTGTCTCTTTAAACTGGA	1019
1000	DB	GGTAGGCTTAATCTCAGAAAGGGCCGTTCAAAGCAGACGCTGTCTCTTTAAACTGGA	1059
1020	QY	TTGCRAAATGGGAAGTCAGTGTATCAGACAAAGCTAGCAGTTGTAGATGAATGGAGAAAAG	1079
1060	DB	TTGCRAAATGGGAAGTCAGTGTATCAGACAAAGCTAGCAGTTGTAGATGAATGGAGAAAAG	1119
1080	QY	AATGGCCCTGTGCTGGAAGAAAGTGGCATATCTTGGAAATGAAGTGTCTGTATGAAGAGTG	1139
1120	DB	AATGGCCCTGTGCTGGAAGAAAGTGGCATATCTTGGAAATGAAGTGTCTGTATGAAGAGTG	1179
1140	QY	CTTGAAGAGATGGGCTTAAGTGGCCCTCTGTGATGCTGTCTAGCCACAGAGGC	1199
1180	DB	CTTGAAGAGATGGGCTTAAGTGGCCCTCTGTGATGCTGTCTAGCCACAGAGGC	1239
1200	QY	TGTTGGATACATTTGCAAAATGTAATGGTGGCCGTGTGCGCATCCGAGAATTTGCAGCA	1259
1240	DB	TGTTGGATACATTTGCAAAATGTAATGGTGGCCGTGTGCGCATCCGAGAATTTGCAGCA	1299
1260	QY	CATTGGCTACTAATGGAAAAAGTTAATAATTCATGCCAAAATAG	1305
1300	DB	CATTGGCTACTAATGGAAAAAGTTAATAATTCATGCCAAAATAG	1345

RESULT	9	BD260951	1772 bp	DNA	linear	PAT 17-JUL-2003
LOCUS						
DEFINITION		Carbohydrate-modifying enzymes.				
ACCESSION		BD260951				
VERSION		BD260951.1	GI:33070721			
KEYWORDS		JP 2002541836-A/1.				
SOURCE		Homo sapiens (human);				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS		1 'bases 1 to 1772)				
TITLE		Lal, P., Yue, H., Tang, T. Y., Hillman, J. L., Baughn, M. R. and Yang, J.				
JOURNAL		Carbohydrate-modifying enzymes				
COMMENT		Patent: JP 2002541836-A 1 10-DEC-2002; INCYTE GENOMICS INC				
		OS Homo sapiens (human)				
		PN JP 2002541836-A/1				
		PD 10-DEC-2002				
		PF 20-APR-2000 JP 2000612430				
		PR 21-APR-1999 US 60/130383				
		PI PREETI LAL, HENRY YUE, TOM Y TANG, JENNIFER L HILLMAN, MARIAH R BAUGHN,				
		PI BAUGHN,				
		PI JUNMING YANG				
		PC C:2N15/09, A61K38/00, A61K38/43, A61K45/00, A61P3/10, A61P29/00, A61P35/00,				
		A61P37/02, C07K16/40, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/00,				
		PC 24, C12Q1/68,				
		PC G01N33/15, G01N33/50, G01N33/53, G01N33/566, G01N33/68, C12N15/00,				
		PC C12N5/00,				
		PC A61K37/02, A61K37/48				
		CC Incyte ID No: 000422CB1				
		Key				
		FT source				
		1..1772				
		/organism='Homo sapiens (human)'				
FEATURES		Location/Qualifiers				
source		1..1772				

Thu May 13 11:53:01 2004

```

ACCESSION      BC063776
VERSION        BC063776.1  GI:39795344
KEYWORDS       Mus musculus (house mouse)
SOURCE         MGC.
ORGANISM       Mus musculus
REFERENCE
AUTHORS        Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,
                Klausner, R.E., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
                Altschul, S.F., Zeeberg, K.H., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
                Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
                Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
                Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
                Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
                Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
                Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J.,
                McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
                Woxley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
                Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
                Pahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
                Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
                Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
                Dickson, V.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
                Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.B.,
                Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
                Generation and initial analysis of more than 15,000 full-length
                human and mouse cDNA sequences
                Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002);
                12477932
                2 (bases 1 to 1704)
                Direct Submission
                Submitted (03-DEC-2003) National Institutes of Health, Mammalian
                Gene Collection (MGC), Cancer Genomics Office, National Cancer
                Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
                USA
                NIH-MGC Project URL: http://mgc.nci.nih.gov
                Contact: MGC help desk
                Email: cgabbs@mail.nih.gov
                Tissue Procurement: Jeffrey E. Green, M.D.
                cDNA Library Preparation: Life Technologies, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Genome Sequence Centre,
                BC Cancer Agency, Vancouver, BC, Canada
                info@bcgsc.bc.ca
                Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
                Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
                Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
                Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
                Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
                Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
                Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Varady,
                George Yang, Scott Zuyderduyn, Marco Marra.
                Clone distribution: MGC clone distribution information can be found
                through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
                Series: IRAK Plate: 131 Row: n Column: 21
                This clone was selected for full length sequencing because it
                passed the following selection criteria: matched mRNA gi: 22238853.
                Location/Qualifiers
                1..1704
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="MGC:70211 IMAGE:6398450"
                /tissue_type="Colon, normal. 5 month old male mouse."
                /clone_lib="NCI CGAP_Co24"
                /lab_host="DH10B"
                /note="vector: pCMV-SPORT6"
                17..1315
                /codon_start=1
FEATURES
source
CDS

```

```

/product="Unknown (protein for MGC:70211)"
/protein_id="AAH63776.1"
/db_xref="GI:39795345"
/translation="MDALEKGAATSGPAPRPSRPPKLSRSGRAGLEKPPHIA
ALVLRGSGKGIPIKIKRAGVPLIGWLRALDAGVQSVWSTDDDEENVAQKQF
CAQVHRSSTSDSSTSLDAIVFLNVEVDIVGNIQATSFCLEHFDIDLVKQAMIR
EAGYDVSFVVRHQFMSSETKQGVRETEPLINLPAKPRPRDWDGDELYGKSYFA
KRHLIEMLYOGKMAVYEMAEHSVDIDVDWPIAQORVLRFGYFGKEKKEIKLL
VKNIDGCLTNGHIVSGDOKETISVDVDAIGISLLKKSIGIEVLISERACSKQTISA
LKNDKTSVSDSKLITVDENWKEMCLCKEVAVLGNESDEECLKRVCLSAVPADAC
SQAQKAGVYICKSGGGRGAIREFAEHIFLLIEKVNNSQK"
146..781
/note="CTP transf 3; Region: Cytidylyltransferase. This
family consists of two main Cytidylyltransferase
activities: 1) 3-deoxy-manno-octulosonate
cytidylyltransferase,, EC:2.7.7.38 catalysing the
reaction: CTP + 3-deoxy-D-manno-octulosonate <=>
diphosphate + CMP-3-deoxy-D-manno-octulosonate, 2)
acetylneuraminate cytidylyltransferase EC:2.7.7.43,,
catalysing the reaction: CTP + N-acetylneuraminate <=>
diphosphate + CMP-N-acetylneuraminate"
/db_xref="CDD:pfam02348"
ORIGIN
Query Match 77.3%; Score 1008.8; DB 10; Length 1704;
Best Local Similarity 87.1%; 2red. No. 6.5e-231;
Matches 1138; Conservative 4; Mismatches 156; Indels 8; Gaps 3;
QY 1 ATGGACTCGGTGGAGAGGGGGCGGCACCTCGTCTCCACCCGGGGGGGACCGTCC 60
DB 17 ATGGAGCGCTGGAGAGGGGGCGGCACCTCGTGGGGGGCGGTGGACGGCGTCC 76
QY 61 CGGGGGCGCGCGCGGAGCTGCGAGCGCACTCTCGCGGGGGCGAGCGCCGAGGTGGAG 120
DB 77 CGGGGGCGCGCGCGGAGCTGCGAGCGCA-----GCGGGGGCGCGGGCGCGCTAGAG 130
QY 121 AAGCCCGCGCACCTGCGCGAGCGCCCTAAATTCTGCGCGGGGGAGCGAGCAAGCAATCCCGCTG 180
DB 131 AAGCCCGCGCACCTGCGAGCGCTGCTGCTGCGCGGGGGCGAGCAAGCAATCCCACTG 190
QY 181 AAGAACATTAAAGCACTGGCGGGGGTCCCGCTCATTTGGCTGGTCTCGCTGGGGCGCTG 240
DB 191 AAGAACATCAAGCGCTTGGCGGGGGTCCCGCTCATTTGGCTGGTCTCGCTGGGGCGCTG 250
QY 241 GATTGAGGGCGCTTCCAGAGTGTATGGTTTCCAGACACCATCATGATGAAATGAGAATGTG 300
DB 251 GATGCGGGGGTCTTCCAGAGTGTGTTGGTTTCCAGACACCATCATGATGAAATGAGAATGTG 310
QY 301 GCCAAACAAATTTGGTGCAAGTTTCATCGAAGAGTTTCTGAAGTTTCAAAGACAGCTCT 360
DB 311 GCCAAACAGTTTGGTGCAAGTTTCATCGAAGAGTTTCTGAAGTTTCAAAGACAGCTCT 370
QY 361 ACTCCTAGTAGCCATCATAGAAATTTCTTAATATATATATGAGGAGTGCATTTGATGGA 420
DB 371 ACCTCCTAGTAGCCATCATAGAAATTTCTTAATATATATGAGGAGTGCATTTGATGGA 430
QY 421 AATATTCAAGCTACTTCTCATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 480
DB 431 AATATCAAGCCCATCTCTCATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 490
QY 481 ATTTCGAGAAGAAGATATGATTTCTGTTTCTTGTGTTGAGAGCGCCATCAGTTTCGATGG 540
DB 491 ATCCGAGAAGAAGATATGATTTCTGTTTCTTGTGTTGAGAGCGCCATCAGTTTCGATGG 550
QY 541 AGTGAATTCAGAAAGAGTTTCGTTGAAAGTGAACCGAACTCTGATTTTAATTCAGCTAAA 600
DB 551 AGTGAATTCAGAAAGAGTTTCGTTGAAAGTGAACCGCTCTGAACTTCAATTCAGTCCGAAA 610
QY 601 CGGCTCTGTGACAGAGTGGAGTGGAGATTTATATGAAATGCTCATTTTATTTTCT 660
DB 611 CGGCTCTGTGACAGAGTGGAGTGGAGATTTATGAAACCGCTCTCATTTTATTTTCT 670
QY 661 AAGAGACATTGTAGAGATGGGTTTACTTTCAGGGT--GGAAATGGCATCTACTAGAAATG 719

```



```

Db      671  AAAAGCACTTGTATGATGATGGTACTTACTACAGGTGGGAAATGCGTATTATGAATG 730
QY      720  CGAGCTGGAACATAGTGTGGATAGATGCGGATATGATGCGCTATTGCGATGAGCAAG 779
Db      731  CGAGCT-GAGCAGTGTGTGATTCGAGTGGACATGATGCGGATGCGGATGCGAGCAAG 789
QY      780  AGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAGCAATATAAACTTTGGTTG 839
Db      790  AGTCTGAGATTTGGCTATTTTGGAAAGAGAGCTGAGGAGATATAAGCTTTGGTTG 849
QY      840  CAATATTGATGATGCTGCTCAATGCGCCCHTTTATGATCAGGACACCAAAAGAAAT 899
Db      850  TAATATTGATGATGCTGCTCAATGCGCCCHTTTATGATCAGGACACCAAAAGAAAT 909
QY      900  AATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 959
Db      910  AATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 969
QY      960  GGTGAGCTTAATCTCAGAAAGGGCTGTTTCAAGCAGACGCTGCTTCTTTAAACTGGA 1019
Db      970  GGTGAGCTCATCTCAGACGGGCTGCTTCCAGCAGACGCTTCCGCGCTTAAAGCTGGA 1029
QY      1020  TCGCAAAATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1079
Db      1030  CTGTAAACAGAGCTAGTGTGTCGATAAGCTGCGCACCGCTGCTGATGATGATGATGAT 1089
QY      1080  AATGGGCTGTGCTGGAAGAGTGGCATATCTTGGAAATGAACTGCTGATGATGATGATGAT 1139
Db      1090  GATGGGCTGTGCTGGAAGAGTGGCATATCTTGGAAATGAACTGCTGATGATGATGATGAT 1149
QY      1140  CTTCAAGAGATGGGCTTAAGTGGCTTCTGATGATGATGATGATGATGATGATGATGATGAT 1199
Db      1150  CTTCAAGAGATGGGCTTAAGTGGCTTCTGATGATGATGATGATGATGATGATGATGATGAT 1209
QY      1200  TGTGGATACATTTGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1259
Db      1210  TGTGGGATACATCTGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1269
QY      1260  CATTTGCTTACTAATGGAAGATTAATAATTCATGCGCAAAATAG 1305
Db      1270  CATTTGCTTACTAATGGAAGATTAATAATTCATGCGCAAAATAG 1315

RESULT 12
LOCUS   BC031500
DEFINITION Mus musculus cytidine monophospho-N-acetylneuraminic acid synthetase, mRNA (cdna clone MGC:27736 IMAGE:2647633), complete cds.
ACCESSION BC031500
VERSION   BC031500.1 GI:2619375
KEYWORDS MGC.
SOURCE    Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1728)
AUTHORS   Strausberg,R.L., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Klausner,R.T., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Altschul,S.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,E., Hopkins,K.J., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Diatchenko,L., Marusina,K., Bonaldo,M.F., Casavant,T.L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Schetz,T.E., Brownstein,M.J., Tsdirn,T.B., Toshiyuki,S., Carinci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abrarson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Guaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,

```

Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Kravinsky,M.I., Skalska,J., Smalhus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marta,M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE
PUBMED 12477932
2 (bases 1 to 1728)
Strausberg,R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT
NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk
Email: mgc@nci.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gumaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAK plate: 35 Row: b Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22208853.

FEATURES
source

1..1728
/organism="Mus musculus"
/mol_type="mRNA"
/strain="PVB/N"
/db_xref="taxon:10090"
/clone="MGC:27736 IMAGE:2647633"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP Mam1"
/lab_host="DHICB"
/note="Vector: pCMV-SPORT6"
1..1728
/gene="Cmas"
/note="synonym: CMP-Neu5Ac"
/db_xref="LocusID:12764"
/db_xref="MGI:1337124"
17..1315
/codon_start=1
/product="cytidine monophospho-N-acetylneuraminic acid synthetase"
/protein_id="AAH31500.1"
/db_xref="GI:21619376"
/db_xref="LocusID:12764"
/translation="MDALEKGSATSPAPRGSRPPKQSRGAGRGKPPHLLA
ALIVARGSGKIPKRIKFLAGVPLIGVIRALDAGVFSQSVVSTDDHEDVAKQF
CAQVRRSSTSKDSTLDAIVEFLNHYNEVDVGNIOATSPCLHTDLOKVAEMIR
EGYDSVFSVRHQFVRSWEIOKGVRETEPLNAPKPRQDQDNGELVNGSFVFA
KRHLIEMGLQGMAYEMRAHSVDIVDIDWPTAEQVRLFRFGFGKEKLEIKLL
VNDIGCLTNGHLYVSGDQKEIISYDVKDAIGSILKISGIEVRLISERACSKQTLGA
LKDDCTEVSVDKATGATDWEKMGCLWKEVLYLGNVSDDECLKRVGLSAVPADAC
SQAQVAGYICKSGRGRAIRFAEHLFLLIEKVNNSQCK"

gene

CDS

misc_feature

146..781
/note="CTP transf 3; Region: Cytidyltransferase. This family consists of two main Cytidyltransferase activities: 1) 3-deoxy-manno-octulosonate Cytidyltransferase, EC:2.7.7.38 catalysing the reaction:- CTP + 3-deoxy-D-manno-octulosonate <=>

```

misc_feature
818..1273
/note="COG1778; Region: Low specificity phosphatase (HAD
superfamily) [General function prediction only];"
/db_xref="CDD:pfam02348"

ORIGIN
Query Match 77.3%; Score 1008.8; DB 10; Length 1728;
Best Local Similarity 87.1%; Pred. No. 6.5e-231;
Matches 1138; Conservative 4; Mismatches 156; Indels 8; Gaps 3;

QY 1 ATGACTCGTGAGAGAGGGGGCGGACCTCGCTCCACACCGCGGGGGCGGACCGTCC 60
Db 17 ATGACGCGCTGAGAGAGGGGGCGGACCTCGCTCGGCGGCGGCGGCGGCGGCGTCC 76
QY 61 CGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 77 CGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 130
QY 121 AAGCCCGCGCACCTGGCAGCCCTCAATTCCTGGCGGGGAGGCGAGCAAGGCAATCCCGCTG 180
Db 131 AAGCCCGCGCACCTGGCAGCCCTCAATTCCTGGCGGGGAGGCGAGCAAGGCAATCCCGCTG 190
QY 181 AAGAACATTAAGCACCTGGCGGGGGTCCCGCTCATTTGGCTGGGTCCTGGCGGGCGGCGGCTG 240
Db 191 AAGAACATTAAGCACCTGGCGGGGGTCCCGCTCATTTGGCTGGGTCCTGGCGGGCGGCGGCTG 250
QY 241 GATTGAGGCGCTTCCAGAGTGTATGGTTTCGACAGACCATGATGAATTTGAGATGTG 300
Db 251 GATTGAGGCGCTTCCAGAGTGTATGGTTTCGACAGACCATGATGAATTTGAGATGTG 310
QY 301 GCCAACAATTTGGTGACAGAGTTCATCGAAGAGTTCGAGTTTCAAGAGTTCGAGAGTCT 360
Db 311 GCCAACAATTTGGTGACAGAGTTCATCGAAGAGTTCGAGTTTCAAGAGTTCGAGAGTCT 370
QY 361 ACCTCACTAGATGCCATCATAGAAATTTTAAATATATATATGAGTGGTGCATTTGATGAG 420
Db 371 ACCTCACTAGATGCCATCATAGAAATTTTAAATATATATATGAGTGGTGCATTTGATGAG 430
QY 421 AATATTCAAGCTACTCTCATGTTTACATCTACTGATCTTCAAAAGTTTCGAGAAATG 480
Db 431 AATATTCAAGCTACTCTCATGTTTACATCTACTGATCTTCAAAAGTTTCGAGAAATG 490
QY 481 ATTGGAAGAGAGATATGATTTCTGTTTCTCTGTTGAGAGCCCATCAGTTTCGATGG 540
Db 491 ATTGGAAGAGAGATATGATTTCTGTTTCTCTGTTGAGAGCCCATCAGTTTCGATGG 550
QY 541 AGTGAATTCGAGAGAGTTCGTAAGTGCAGCAACCTCTGAAATTTAAATCCAGCTAAA 600
Db 551 AGTGAATTCGAGAGAGTTCGTAAGTGCAGCAACCTCTGAAATTTAAATCCAGCTAAA 610
QY 601 CGGCTCTGTCGACAAAGCTGGATGAGAGATATATGAGAAATGCTCATTTATTTGCT 660
Db 611 CGGCTCTGTCGACAAAGCTGGATGAGAGATATATGAGAAATGCTCATTTATTTGCT 670
QY 661 AAAAGACATTTGATAGATGGTTTACTTGCAGGCTTGGAGATGATGATGAGTTCGAGCAAG 719
Db 671 AAAAGACATTTGATAGATGGTTTACTTGCAGGCTTGGAGATGATGATGAGTTCGAGCAAG 730
QY 720 CGAGCTCGAACATAGTGTGGATATAGATGATGATGATGATGATGATGATGATGATGATG 779
Db 731 CGAGCTCGAACATAGTGTGGATATAGATGATGATGATGATGATGATGATGATGATGATG 789
QY 780 AGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAAGAAATATAAATCTTTGGTTG 839
Db 790 AGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAAGAAATATAAATCTTTGGTTG 849
QY 840 CAATATTGATGATGCTTCCCAATGGCCCATTTTGTATGATGATGATGATGATGATGATGAT 899

```

```

Db 850 TAAATATTGATGGAGTCTCTACCAATGGCCACATTTATGATGATGATGATGATGATGATGAT 909
QY 900 AATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 959
Db 910 AATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 969
QY 960 GGTGAGGCTTATCTCAGAAAGGGCGCTGTTTCAAGCAGACGCTGCTCTTTTAAACCTGGA 1019
Db 970 GGTGAGGCTTATCTCAGAAAGGGCGCTGTTTCAAGCAGACGCTGCTCTTTTAAACCTGGA 1029
QY 1020 TTGCAAAATGGAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1079
Db 1030 CTGTAAAAACAGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1089
QY 1080 AATGGGCGCTGCTGGAAGAGAGTGCATATCTTGGAAATGAAGTGTCTGATGATGATGATG 1139
Db 1090 GATGGGCGCTGCTGGAAGAGAGTGCATATCTTGGAAATGAAGTGTCTGATGATGATGATG 1149
QY 1140 CTTGAGAGAGTGGGCTTAAGTGGCGCTGCTGCTGATGATGATGATGATGATGATGATGATGATG 1199
Db 1150 CTTGAGAGAGTGGGCTTAAGTGGCGCTGCTGCTGATGATGATGATGATGATGATGATGATGATG 1209
QY 1200 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1259
Db 1210 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1269
QY 1260 CATTTGCTTACTAATGCAAAAGTAAATATATGATGATGATGATGATGATGATGATGATGATG 1305
Db 1270 CATTTGCTTACTAATGCAAAAGTAAATATATGATGATGATGATGATGATGATGATGATGATG 1315

```

```

RESULT 13
MMU6215
LOCUS Mus musculus mRNA for CMP-N-acetylneuraminic acid synthetase. 1687 bp mRNA linear ROD 07-AUG-1998
DEFINITION Mus musculus mRNA for CMP-N-acetylneuraminic acid synthetase.
ACCESSION AJ006215.1 GI:3413319
VERSION AJ006215.1
KEYWORDS CMP-N-acetylneuraminic acid synthetase.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 Munster, A.K., Eckhardt, M., Potvin, B., Mühlenhoff, M., Stanley, P. and
Gerardy-Schahn, R.
Mammalian cytidine 5'-monophosphate N-acetylneuraminic acid
synthetase: a nuclear protein with evolutionarily conserved
structural motifs
Proc. Natl. Acad. Sci. U.S.A. 95 (16), 9140-9145 (1998)
JOURNAL 98356124
MEDLINE 9689047
PUBMED
REFERENCE 2 (bases 1 to 1687)
Gerardy-Schahn, R.
Direct Submission
AUTHORS Submitted (13-MAY-1998) Gerardy-Schahn R., Institut fuer
Medizinische Mikrobiologie, Medizinische Hochschule Hannover,
Carl-Neuberg-Str. 1, Hannover, 30625, GERMANY
JOURNAL Location/Qualifiers
FEATURES
source
1..1687
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/cell_line="AtT-20"
1..1299
/EC_number="2.7.4.3"
/codon_start=1
/product="CMP-N-acetylneuraminic acid synthetase"
/protein_id="CAA06915.1"
/db_xref="GI:3413320"
/db_xref="GOA:O88719"
/db_xref="SPTREMBL:O88719"
/translaton="MDALEKAVTSGPAPRPSRPPKQSRGAGRGLEKPEHLA
ALVLRAGSGKIPKTKRLAGVPLIGWLEALDAGVFQVWVSTDHDEIENVAQF

```


REMARK

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT

Contact: MGC help desk
Email: cgabs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunsaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lousseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.K., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 3 Row: b Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22027483.

FEATURES

source
1..1588
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:5424 IMAGE:3448277"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_10"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
1..1588
/gene="CMAS"
/db_xref="LocusID:55907"
/db_xref="MIM:603316"
80..871
/codon_start=1
/product="CMAS protein"
/protein_id="AAH16609.1"
/db_xref="GI:116741605"
/db_xref="LocusID:55907"
/translation="MDSVERGATSVNPRGSRPRPKLQNSGSGQGRGVEKPH
LAALLARGSKGPIPLNKLHLAGVPLGLVLRALDSGAFQSVWSDHDEIENVAK
QFGAQRHRSSEVSDSLDILFNLHNEVDVGNQATSPCLHPDLOKVAEM
IREGVSVFVWRHGFVSEIQRVTEPDLNPAKPRQDWDGDIENGFSY
FAKHLLJEMVLOGGKAYEMRAHSVDIDVDIDWPIAEQRVLR"
215..850
/note="CTP trans 3; Region: Cytidylyltransferase. This family consists of two main Cytidylyltransferase activities: 1) 3-deoxy-manno-octulosonate cytidylyltransferase, EC:2.7.7.38 catalysing the reaction: CTP + 3-deoxy-D-manno-octulosonate <=> diphosphate + CTP + 3-deoxy-D-manno-octulosonate, 2) acylneuraminate cytidylyltransferase EC:2.7.7.43, catalysing the reaction: CTP + N-acylneuraminate <=> diphosphate + CTP + N-acylneuraminate"
/db_xref="CDD:pfam02348"

misc_feature

215..850
/note="CTP trans 3; Region: Cytidylyltransferase. This family consists of two main Cytidylyltransferase activities: 1) 3-deoxy-manno-octulosonate cytidylyltransferase, EC:2.7.7.38 catalysing the reaction: CTP + 3-deoxy-D-manno-octulosonate <=> diphosphate + CTP + 3-deoxy-D-manno-octulosonate, 2) acylneuraminate cytidylyltransferase EC:2.7.7.43, catalysing the reaction: CTP + N-acylneuraminate <=> diphosphate + CTP + N-acylneuraminate"
/db_xref="CDD:pfam02348"

ORIGIN

Query Match 72.4%; Score 918.4; DB 9; Length 1588;
Best Local Similarity 86.8%; Pred No. 3e-209;
Matches 1123; Conservative 4; Mismatches 5; Indels 174; Gaps 3;
QY 1 ATGGAATCGGTGGAGAGAGGGGCGCCACCTCGCTCCAAACCCCGGGGGGCGCGTCC 60
DB 80 ATGGAATCGGTGGAGAGAGGGGCGCCACCTCGCTCCAAACCCCGGGGGGCGCGTCC 139
QY 61 CGGGGCGGGCGCCGGAAGCTGCGGCACTCTCGCGGCGCGGCGGCGGCGGTGGAG 120
DB 140 CGGGGCGGGCGCCGGAAGCTGCGGCACTCTCGCGGCGCGGCGGCGGCGGTGGAG 199
QY 121 AAGCCCGCGACCTCGGAGCCCTAATCTGGCCCGGGGAGGCGAGCAAGGCAATCCCGCTG 180
DB 200 AAGCCCGCGACCTCGGAGCCCTAATCTGGCCCGGGGAGGCGAGCAAGGCAATCCCGCTG 259

QY 181 AGAACAATTAAGCACTTGGCGGGGGTCCCGCTCATTTGGTGGTGGTGGTGGCGGCGCTG 240
DB 260 AAGAACATTAAGCACTTGGCGGGGGTCCCGCTCATTTGGTGGTGGTGGTGGCGGCGCTG 319
QY 241 GATTCAGGGGCGCTTCCAGAGTGTATGGGTTTCAGACAGACCATGATGAAATGAGAATGTG 300
DB 320 GATTCAGGGGCGCTTCCAGAGTGTATGGGTTTCAGACAGACCATGATGAAATGAGAATGTG 379
QY 301 GCCAAACAATTTGGTGCACAAAGTTCATCGAAGAAGTTCTGAAGTTTCAAAAGACAGCTCT 360
DB 380 GCCAAACAATTTGGTGCACAAAGTTCATCGAAGAAGTTCTGAAGTTTCAAAAGACAGCTCT 439
QY 361 ACCTCACTAGATGCCATCATAGATTTCTTAATTAATATAGAGTGAAGTGAAGTGAAGTGA 420
DB 440 ACCTCACTAGATGCCATCATAGATTTCTTAATTAATATAGAGTGAAGTGAAGTGAAGTGA 499
QY 421 AATATTCGAAGTACTTCTTCATGTTTATACATCTACTGATCTTCAAAAGTTTGCAGAAATG 480
DB 500 AATATTCGAAGTACTTCTTCATGTTTATACATCTACTGATCTTCAAAAGTTTGCAGAAATG 559
QY 481 ATTGGAAGAAGGATATGATTTCTGKTTTCTGTTTGTGAGACCGCATCAGTTTTCGATGG 540
DB 560 ATTGGAAGAAGGATATGATTTCTGKTTTCTGTTTGTGAGACCGCATCAGTTTTCGATGG 619
QY 541 AGTGAAATTCAGAAAGGAGTTCTGTAAGTGAACCAACCTCTGAATTTAAATCCAGCTAAA 600
DB 620 AGTGAAATTCAGAAAGGAGTTCTGTAAGTGAACCAACCTCTGAATTTAAATCCAGCTAAA 679
QY 601 CGGCTCTGTCGACAAAGACTGGGATGGAGAAATATATGAAATGGCTCAATTTATTTTCT 660
DB 680 CGGCTCTGTCGACAAAGACTGGGATGGAGAAATATATGAAATGGCTCAATTTATTTTCT 739
QY 661 AAAAGACATTTGATAGAGATGGGTTTACTTGGAGGGTGGAAATGGCATCTACGAAATG 719
DB 740 AAAAGACATTTGATAGAGATGGGTTTACTTGGAGGGTGGAAATGGCATCTACGAAATG 799
QY 720 CGAGCTCGAACATAGTGGGATATAGATGTCGATTTGATTTGGCTTATGAGACCAAG 779
DB 800 CGAGCT-GAACATAGTGGGATATAGATGTCGATTTGATTTGGCTTATGAGACCAAG 858
QY 780 AGTATTAGATATGCTATTTTGGCAAGAGAGCTTAAGGAAATATAAACTTTTGGTTG 839
DB 859 AGTATTATTA----- 865
QY 840 CAATATTGATGGATGTCACCAATGGCCACATTTATGATATCAGGAGACCAAAAGAAAT 899
DB 866 ----- 865
QY 900 AATATCTTATGATGTAAAGATGCTATTGGGATAGTTTATTAAAGAAAGTGGTATTGA 959
DB 866 -----A 866
QY 960 GGTGAGGCTAATCTCAGAAAGGGGCTGTTCAAAGCAGACGCTGTCTTTTAAACTGGA 1019
DB 867 GGTGAGGCTAATCTCAGAAAGGGGCTGTTCAAAGCAGACGCTGTCTTTTAAACTGGA 926
QY 1020 TTGCAAAATGGAAGTCACTGTATCAGACAGCTAGCTGTAGATGATGATGAGAAAGA 1079
DB 927 TTGCAAAATGGAAGTCACTGTATCAGACAGCTAGCTGTAGATGATGAGAAAGA 986
QY 1080 AATGGGCGCTGTGCTGGAAGAAGTGGCATATCTTGGAAATGAAGTGTCTGTGAAAGAGTG 1139
DB 987 AATGGGCGCTGTGCTGGAAGAAGTGGCATATCTTGGAAATGAAGTGTCTGTGAAAGAGTG 1046
QY 1140 CTTTGAAGAGAGTGGGCGCTTAAGTGGCGCTCTCTGATAGCTCTTCTCAGCCCGCAAGAGC 1199
DB 1047 CTTTGAAGAGAGTGGGCGCTTAAGTGGCGCTCTCTGATAGCTCTTCTCAGCCCGCAAGAGC 1106
QY 1200 TGTTCGATACATTTGCAATATGATGGTGGCGGTGGTCCATCCGAGATTTTCAGAGCA 1259
DB 1107 TGTTCGATACATTTGCAATATGATGGTGGCGGTGGTCCATCCGAGATTTTCAGAGCA 1166
QY 1260 CATTTTGGCTACTACTTAAGGAAAGATTAATAATTCATGCCAAATAATAG 1305

DB 1167 CATTGGCTACTAATGGAAAGGTTAATAATCATGCCAAAATAG 1212

RESULT 15

AX541147

LOCUS

AX541147

DEFINITION

Sequence 394 from Patent WO20055700.

ACCESSION

AX541147

VERSION

AX541147.1 GI:25274513

KEYWORDS

Homo sapiens (human)

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

Escobedo, J., Garcia, P.D., Kassam, A., Lamson, G., Drmanac, R.,

Crkvenjakov, R., Dickson, M., Drmanac, S., Labat, I., Leshkowitz, D.,

Kita, D., Garcia, V., Jones, W.L., Stache-Crain, B. and Scott, E.M.

Human genes and gene expression products isolated from human

prostate

Patent: WO 0205700-A 394 18-JUL-2002;

Chiron Corporation (US); HYSEQ, INC. (US)

FEATURES

Location/Qualifiers

1..838

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match

Best Local Similarity

Matches 715; Conservative 4; Mismatches 26; Indels 10; Gaps 6;

48.8%; Score 636.6; DB 6; Length 838;

94.7%; Pred. No. 1e-141;

150 GGCCGGGGGAGGAGCAAGGATCCCTCGAGAAACATTAAACACCTGGCGGGGTCCC 209

DB 71 GGCCGGGGNAGGAGCAANGCATCCCTCGAGAAACATTAAACACCTGGCGGGGTCCC 130

210 GTCATTGGCTGGGTCTGGTGGCGGCTGGATTGAGGAGGCTTCCAGAGTGTATGGGT 269

DB 131 GTCATTGGCTGGGTCTGGTGGCGGCTGGATTGAGGAGGCTTCCAGAGTGTATGGGT 190

270 TTCGACAGACCATGATGAATTCAGAAATGTCGCAACAACTTTGGTGCACAGTTCATCG 329

DB 191 TTCGACAGACCATGATGAATTCAGAAATGTCGCAACAACTTTGGTGCACAGTTCATCG 250

330 AAGAGTCTCTGAAGTTTCAAAGACAGCTCTACCTCACTAGATGCCATCATAGAAATTC 389

DB 251 AAGAGTCTCTGAAGTTTCAAAGACAGCTCTACCTCACTAGATGCCATCATAGAAATTC 310

390 TAATATATATAGAGGTGACATTTAGGAATATTAAGTACTCTCTCTCTCTCTCTCTCT 449

DB 311 TAATATATATAGAGGTGACATTTAGGAATATTAAGTACTCTCTCTCTCTCTCTCTCT 370

450 TCT 509

DB 371 TCT 430

510 CTCTGTGTGAGAGCGGCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTCGTGAAGT 569

DB 431 CTCTGTGTGAGAGCGGCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTCGTGAAGT 490

570 GACCGACCTCTGAATTAATTCAGCTTAACGCGCTCTCGACACAGCTGGGATGGAGA 629

DB 491 GACCGACCTCTGAATTAATTCAGCTTAACGCGCTCTCGACACAGCTGGGATGGAGA 550

630 ATTATATGAATGGGTCAATTTATTTGCTAAAGACATTTGATAGAGATGGGTACIT 689

DB 551 ATTATATGAATGGGTCAATTTATTTGCTAAAGACATTTGATAGAGATGGGTACIT 610

690 GCAGGGTGG-AAAATGGCTACTACGAAATCGAGCTGGAACTAGTGTGGATATAGATG 748

DB 611 GCAGGGTGGAAATGGCTACTACGAAATCGAGCTGGAACTAGTGTGGATATAGATG 669

QY 749 TCGATATTGATTGGCTATTTCAGAGCAAGAGTATTAGATATGCTATTTTGGCAAG 808
DB 670 TCGATATTGATTGGCTATTTCAGAGCAAGAGTATTAGATATGCTATTTTGG-TRAG 728
QY 809 AGAGCTTAAGCAAAATAAAACTTTTGGTTTGAATATTGATGATGCTCTCACCATGGCC 868
DB 729 AAGAGCTTAAG--AATAAACTTTTGGTT--GCATATTGATGATGCTCTCACC- 781
QY 869 ACATTATGATATCAGGAGACCAAAAGAAATAATA 903
DB 782 CCNTTTTCTGTCAGGAGACCAAAAAA 816

Search completed: May 11, 2004, 17:15:27

Job time : 5315.32 secs

Sequence 1, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 18033, A
Sequence 28, Appl
Sequence 1892, Ap
Sequence 1282, Ap
Sequence 1, Appli
Sequence 14795, A
Sequence 398, App
Sequence 14544, A
Sequence 14820, A
Sequence 5, Appli
Sequence 5, Appli
Sequence 1, Appli
Sequence 4468, Ap
Sequence 16, Appl
Sequence 2696, Ap

ALIGNMENTS

RESULT 1
US-09-516-143A-1
; Sequence 1, Application US/09516143A
; Patent No. 6333182
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PF505PCT
; CURRENT APPLICATION NUMBER: US/09/516,143A
; PRIOR FILING DATE: 2003-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 1
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: {1}...(1302)
US-09-516-143A-1

Query Match 99.9%; Score 1303.4; DB 4; Length 1305;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGACTCGGTGGAGAAAGGGGGCGCCACCTTCCTCCAAACCCCGGGGGCGACCGTCC	60
Db	1	ATGACTCGGTGGAGAAAGGGGGCGCCACCTTCCTCCAAACCCCGGGGGCGACCGTCC	60
Qy	61	CGGGGCGGGCGCGGAGCTGCAGGGAATCTCGCGCGGGCGGCGGCGGCTGGAG	120
Db	61	CGGGGCGGGCGCGGAGCTGCAGGGAATCTCGCGCGGGCGGCGGCGGCTGGAG	120
Qy	121	AAGCCCGCGACCTGGCAGCCCTTAATCTGGCCCGGGGAGGAGCAAAAGGCTATCCCTTG	180
Db	121	AAGCCCGCGACCTGGCAGCCCTTAATCTGGCCCGGGGAGGAGCAAAAGGCTATCCCTTG	180
Qy	181	AAGAACATTAAAGACCTGGCGGGGGTCCCGTCTATGGCTGGGCTCGTGGGCGCTG	240
Db	181	AAGAACATTAAAGACCTGGCGGGGGTCCCGTCTATGGCTGGGCTCGTGGGCGCTG	240
Qy	241	GATTTCAGGGCGCTTCAGAGTGTATGGTTTCGACAGACCATGATGAATTCAGAAATG	300
Db	241	GATTTCAGGGCGCTTCAGAGTGTATGGTTTCGACAGACCATGATGAATTCAGAAATG	300
Qy	301	GCCAAACATTTGGTGCAAGTTCATCGAAGAGTTCGAAAGTTTCAAAGACAGCTCT	360
Db	301	GCCAAACATTTGGTGCAAGTTCATCGAAGAGTTCGAAAGTTTCAAAGACAGCTCT	360

US-09-930-440B-3
Perfect score: 1305
Sequence: 1 atggactcgtggagagg.....ataattcagccaaaatag 1305
Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0
Searched: 682709 seqs, 277475446 residues
Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Query Match	Length DB ID		
1	1303.4	99.9	1305	US-09-516-143A-1	Sequence 1, Appli
2	1272.4	97.5	1770	US-09-620-312D-360	Sequence 360, App
3	53.6	4.9	7218	US-08-232-463-14	Sequence 14, Appli
4	50.6	3.9	1830121	US-09-557-884-1	Sequence 1, Appli
5	50.6	3.9	1830121	US-09-543-990A-1	Sequence 1, Appli
6	50.2	3.8	585	US-09-543-681A-1730	Sequence 1730, Ap
7	45.8	3.5	580073	US-08-545-528D-1	Sequence 1, Appli
8	45.6	3.5	684	PCT-US96-05320A-1282	Sequence 1282, Ap
9	42.4	3.2	832	US-09-621-976-2813	Sequence 2813, Ap
10	41.6	3.2	113211	US-09-596-002-40	Sequence 40, Appl
11	41	3.1	1182	US-09-252-991A-6736	Sequence 6736, Ap
12	41	3.1	1848	US-09-252-991A-6953	Sequence 6953, Ap
13	41	3.1	1899	US-09-252-991A-6781	Sequence 6781, Ap
14	40.4	3.1	744	US-09-252-991A-13301	Sequence 13301, A
15	40.4	3.1	783	US-09-252-991A-12981	Sequence 12981, A
16	40.4	3.1	786	US-09-252-991A-12498	Sequence 12498, A
17	40	3.1	640681	US-09-790-988-1	Sequence 1, Appli
18	39.8	3.0	534	US-09-540-236-1677	Sequence 1677, Ap
19	38.6	3.0	3684	US-08-252-991A-65	Sequence 65, Appl
20	38.6	3.0	7527	US-08-252-991A-71	Sequence 71, Appl
21	38.4	2.9	5652	US-09-601-198-75	Sequence 75, Appl
22	38.2	2.9	1134	US-09-252-991A-8176	Sequence 8176, Ap
23	38.2	2.9	1664976	US-08-916-421B-1	Sequence 1, Appli
24	37.8	2.9	1674	US-08-894-844-14	Sequence 14, Appl
25	37.8	2.9	1674	US-08-894-844-15	Sequence 15, Appl
26	37.8	2.9	4411529	US-09-103-840A-1	Sequence 1, Appli
27	37.6	2.9	468	US-09-134-001C-180	Sequence 180, App

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: C8/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 3.9%; Score 50.6; DB 4; Length 1830121;
Best Local Similarity 45.1%; Pred. No. 0.0098;
Matches 228; Conservative 0; Mismatches 274; Indels 3; Gaps 1;
QY 780 AGTATTAAAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGMAATAAACTTTTGGTTG 839
DB 1747282 ATTTCAAGCTTAAGGAATCCGATGCAACAAAATAGAAAATATAATTTGTCAATAC 1747341
QY 840 CAATATTGATGGATGCTCCACCAATGCCCAATTTATGATATCGAGACCAAAAGAAAT 899
DB 1747342 TGAATGAGCGGCTACTGATGGAACACTTCATATGATGCCAATGGTGAAGCCAT 1747401
QY 900 AATATCTTATGATGTAAAGATGCTATTTGGGATAAGTTTATTAAAGAAAAGTGTATTGA 959
DB 1747402 CAAAAGCTTTCACTAGCAGATGTTTAGGCATATAAAATGCTGATGGATCGGATATCA 1747461
QY 960 GTGAGGCTAATCTCAGAAAGGCTGTTCAAGAGCAGACGCTGTCTTCTTTAAACTGGA 1019
DB 1747462 AGTGGCAGTCTCTTGGTGGCGACTCCCTATTTTACGTGCTGTCGTCATCCGATCTGG 1747521
QY 1020 TTGCAAAATGGAAGTCACTGATATCA---GCAAGCTAGCAGTGTACATGAATGAGAAA 1076
DB 1747522 TATTAATATTCTTCTTGGCAACTTGAAGAAGAACGCTGTGTTGATCTCATGAA 1747581
QY 1077 AGAAATGGGCTGCTGCTGGAAGAGAGTGGGATATCTTTGAAAATGAACTGTCTGATGA 1136
DB 1747582 ACAAGCTGGGCTACCCCGCAGAAACGCTTATATGGGATGATAGCGTAGATCTCC 1747641
QY 1137 GTGCTTCAAGAGAGTGGGCTAGTGGGCTCTGCTGATGCTGCTGCTTCTTACGCCGAAA 1196
DB 1747642 CGCTTGGAGCGGTGGAATCTTTTCTGCTGCTGATGCTGCTGCTTATTTATGTAAGAA 1747701
QY 1197 GGCTGTTGGATACATTTGCAAAATGTAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGAGA 1256
DB 1747702 TGCTGTGATCATGATCTTCCACCCATGCGGCAAGAGGCGCATCTCCGTGAATGTCGA 1747761
QY 1257 GCACATTTGCTACTAATGAAAAA 128-

DB 1747762 TATGATTTTACAGCACAGGAAAA 1747786
RESULT 5
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: the Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: C8/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: C8/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 3.9%; Score 50.6; DB 4; Length 1830121;
Best Local Similarity 45.1%; Pred. No. 0.0098;
Matches 228; Conservative 0; Mismatches 274; Indels 3; Gaps 1;
QY 780 AGTATTAAAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGMAATAAACTTTTGGTTG 839
DB 1747282 ATTTCAAGCTTAAGGAATCCGATGCAACAAAATAGAAAATATAATTTGTCAATAC 1747341
QY 840 CAATATTGATGGATGCTCCACCAATGCCCAATTTATGATATCGAGACCAAAAGAAAT 899
DB 1747342 TGAATGAGCGGCTACTGATGGAACACTTCATATGATGCCAATGGTGAAGCCAT 1747401
QY 900 AATATCTTATGATGTAAAGATGCTATTTGGGATAAGTTTATTAAAGAAAAGTGTATTGA 959
DB 1747402 CAAAAGCTTTCACTAGCAGATGTTTAGGCATATAAAATGCTGATGGATCGGATATCA 1747461
QY 960 GTGAGGCTAATCTCAGAAAGGCTGTTCAAGAGCAGACGCTGTCTTCTTTAAACTGGA 1019

Db 1747462 AGTGGCAGTCTTCTGCGGACATCCCTATTTTACGTGCGCATTCGCGATCTTGG 1747521
 QY 1020 TTGCAAAATGGAAGTCAAGTATCA---CACAGCTAGCAGTGTAGATGAATGAGAAA 1076
 Db 1747522 TATTAAATTTCTTCTTGGCAACTTGAAAGAAACGGCTTGTGTTGATCTCATGA 1747581
 QY 1077 AGAATGGCCCTGTGCTGGAAGAAGTGCATATCTTGAAGTGAAGTGTCTGTGAAGA 1136
 Db 1747582 ACAAGTGGCTCAACGCGCAGCAACCGCTTATATTGGCATGATAGCTAGATCC 1747641
 QY 1137 GTGCTTGAAGAGTGGGCTAAGTGGGCTCCTGCTGATGCTGTCTCTAGCCAGAA 1196
 Db 1747642 CGCCTTTGAGCGTGTGAAGTCTTTTCTGTGGCTGATGCCCTTATTATGTGAAA 1747701
 QY 1197 GCGTGTGTGATACATTTGCAAAATGTAATGTTGGCCGTGGTGGCCATCCGAGAAATTCAGA 1256
 Db 1747702 TGCTGTGTGATCACTTCTTCCACCCTGCGGCAAGGCGCATTCGGTGAATGTCGA 1747761
 QY 1257 GCACATTTGCTTACTTAATGGAATA 1281
 Db 1747762 TATGATTTTACAGGCACAGGAAAA 1747786

RESULT 6

US-09-543-681A-1730
 ; Sequence 1730, Application US/09543681A
 ; Patent No. 6605789
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; PRIOR FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 1730
 ; LENGTH: 585
 ; TYPE: DNA
 ; ORGANISM: Proteus mirabilis
 US-09-543-681A-1730

Query Match 3.8%; Score 50.2; DB 4; Length 585;
 Best Local Similarity 47.8%; Pred. No. 5.2e-05;
 Matches 207; Conservative 0; Mismatches 223; Indels 3; Gaps 2;
 QY 821 AAATAAACTTTTGGTTTCAATATTTGATGATGATGCTCACCAGTCCACATTTATGTAT 880
 Db 92 AAGTCCAATTTGCTGATTTGCGATTTGATGGTGTGATGCTGACGACATTTATATGG 151
 QY 881 CAGGAGACCAAAAGAAATAATATCTTATGATGTAAGATGCTATTGGGATAAGTTTAT 940
 Db 152 GCAATAACGGTGAAGAGTTTGAAGCGTTTAAATGTCGCTGATGCTATGCAATCGTTGT 211
 QY 941 TAAAGAAAGTGGTATTTAGGTTAGGCTAACTCTCAGAAAGGCGCTGTTCAAAGCAGAC 1000
 Db 212 TACTTACATCGGTTATTCAGGTCGCTATCATACCGGTCGCTCAATCTAAGCTTTTAGAAG 271
 QY 1001 TGTCTTTTAAACT--GGATTGCAAAATGGAAGTCACTGTATCA-GACNAGCTAGCAG 1057
 Db 272 ATAGAGCCAAAACACTTGGCATTTACATATCTTTACAGGACAGCAATAAATAGCTTTGG 331
 QY 1058 TTGTAGATGAATGGAAGAAAGAAATGGGCTGCTGCTGGAAGAAGTGGCATATCTTTGAA 1117
 Db 332 CGTCAACAACTGTTAGATACATAAACTTAAACCAAGACAGACAGCTATATTGGCG 391
 QY 1118 ATGAAGTGTCTGATGAAGTGTCTTGAAGAGTGGCGCTTAAGTGGCGCTCTGCTGATG 1177
 Db 392 ATGATCTGATTTGATTTACCGGTAAATGGAAAAATAGGACTGTGAGTTGCCGTTGCTGATG 451
 QY 1178 CTTGTTCTTACGCCAGAGAGGCTGTTGGATACATTTGCAATGTAATGTAATGGTGGCGCTG 1237

Db 452 CCATCCATCTACTCACACCTCGTGGCCGACTATCTTACCCATATTTAGGTGAGCGGTG 511
 QY 1238 CCATCCGAGAAAT 1250
 Db 512 CAGTACGAGAAAT 524

RESULT 7

US-08-545-528D-1
 ; Sequence 1, Application US/08545528D
 ; Patent No. 6537773
 ; GENERAL INFORMATION:
 ; APPLICANT: Fraser et al.
 ; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
 ; Patent No. 6537773
 ; FILE REFERENCE: PB133P1
 ; CURRENT APPLICATION NUMBER: US/08/545,528D
 ; CURRENT FILING DATE: 1995-10-19
 ; PRIOR APPLICATION NUMBER: US 08/488,018
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: US 08/473,545
 ; PRIOR FILING DATE: 1995-06-07
 ; NUMBER OF SEQ ID NOS: 1
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 580073
 ; TYPE: DNA
 ; ORGANISM: Mycoplasma genitalium
 US-08-545-528D-1

Query Match 3.5%; Score 45.8; DB 4; Length 580073;
 Best Local Similarity 48.9%; Pred. No. 0.13;
 Matches 113; Conservative 3; Mismatches 115; Indels 0; Gaps 0;
 QY 370 GATGCCATCATGAAATTTCTTAATATATATGAGGKTCACATTTGAGGAATTTCAA 429
 Db 564078 GAATCATGATTTATGATATCAATGAAAGCTAAAGTTGGTATGTTGATTTCTCAA 564137
 QY 430 GCTACTTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTGACAGAAATGATTCGAGAA 489
 Db 564138 GATGCTTCTTCATCATGCCCAATTTTGGCAAGTAGTTGCTGGAGATACAGTCAATTA 564197
 QY 490 GAAGGATATGATCT 549
 Db 564198 AAGATTTATGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 564257
 QY 550 CAGAAAGGAGTTCGTAAGTGCAGGACCTCTGAATTTAAATCCAGCTAAA 600
 Db 564258 AGGTTGACGCTACTTGAAGGCAATGTAGAAATTTATGAAAGCAATGCCAAA 564308

RESULT 8

PCT-US96-05320A-1282
 ; Sequence 1282, Application PCT/US9605320A
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences
 ; APPLICANT: 9410 Key West Avenue
 ; APPLICANT: Rockville, MD 20850
 ; APPLICANT: United States of America
 ; APPLICANT: Johns Hopkins University
 ; APPLICANT: 720 Rutland Avenue
 ; APPLICANT: Baltimore, MD 21205
 ; APPLICANT: United States of America
 ; APPLICANT: Mark D. Adams
 ; APPLICANT: Owen White
 ; APPLICANT: Hamilton O. Smith
 ; APPLICANT: J. Craig Venter
 ; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
 ; NUMBER OF SEQUENCES: 48
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 ; STREET: 1100 New York Avenue, Suite 600

; PUBLICATION INFORMATION:

US-09-596-002-40

Query Match 3.2%; Score 41.6; DB 4; Length 119211;

Best Local Similarity 44.6%; Pred. No. 0.79;

Matches 207; Conservative 0; Mismatches 254; Indels 3; Gaps 1;

QY 804 CAAGAGAGCTTAAGGAAATAAACTTTTGGTTGCAATATATGATGATGCTCACCAG 863
DB 8025 CTACCAAAAGCCAAACATATCAAACTTTTGGCCATGATGCTGATTTTTCGGA 80084

QY 864 TGSCACATTTATGATCAGGAGACCAAAAGAAATAATCTATGATGATGATGATG 923
DB 80385 TGGCAATCATTTATAATCTGAAGGCACAAACCAACCAATTTATGTCAGATGG 80144

QY 924 TATGGGTAAGTTTATTAAAGAAAGTGGTATTGAGGTGAGGCTAATCTCAGAAAGGC 983
DB 80145 TTTGGGCTTGAAGCGCTAAACAATCAGGCATTATTTGGCCATTATCAGAGTGAAG 80264

QY 984 CTGTTCAAAGCAGCGCTGCTCTTTTAAAGTGG---ATTGCAAAATGGAAGTCAAGTGT 1040
DB 80205 CTCGCCATGTTGACAGACGCGCCAAAGAGCTTGGCATTTCACATATTATCAAGGTCA 80264

QY 1041 ATCAGACAGCTAGCAGTTGATGATGATGGAAGAAAGAAATGGGCTGTGCTGGAAGA 1100
DB 80265 AGATGATAATGACTGTTGGTGGGCTTCTAAGAGCTTGGCATTGAGTTATCACA 80324

QY 1101 AGTGCCATATCTGAAATGAAGTCTGATGAGAGAGTCTTTGAAGAGAGTGGGCTTAAG 1160
DB 80325 CTGTGCTTATATCGGTGATGATTTGCTGATCTAAAGCTGTGCTGAGGAGGTTGG 80384

QY 1161 TGGGCTCTCTGTGATGCTGTCTTCTACCCAGAGAGCTGTGATACATTTGCAATG 1220
DB 80385 GATCAGTTACTAATAGGTGTGAGCAGACCCGAGCAGTTTTCAGATTATATCAACCAA 80444

QY 1221 TAATGGTGGCTGTGCTGATCCAGAAATTTTCAGAGCACATTT 1264

DB 80445 AACTGGGSGTATGTTGCTGTGAGGAGGTTTGTGAGCTGATTT 80498

RESULT 11

US-09-252-991A-6736

; Sequence 6736, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6736

; LENGTH: 1182

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6736

Query Match

Best Local Similarity 3.1%; Score 41; DB 4; Length 1182;

Matches 100; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

QY 46 CGGGGGCGACCTCCGGGCGCGCGCGGAGCTGCAGCGCACTCTCGCGCGGCCAG 105

DB 106 CTGGTGGCGCCCGCCACGGCTGGCCACCGCGGCTATCGGCAAGCTGCTGGTGGCGGTGA 165

QY 106 GCGCGAGGTGTG-GAGAAAGCCCCSCACCTGGCAGCCCTAATTC-TGGCCCGGGAGGCGAG 164

DB 166 GGAACAGCGCGAGAGGCTCCAGGCTTGGCGCCACCGACCGACGCTGGCGGTGCA 225

QY 165 CAAGGCGATCCCTCTGAAGAACATTTAAGCACCTGGGGGGGGTCCCGCTCATTTGGCTGGGT 224
DB 226 CGACCTGATGCTCGCTGGAGCGGCCACCGCTGGCGGGCTTGAGAGCCCCCAGGTCCGCGGGT 285
QY 225 C 225
DB 286 C 286

RESULT 12

US-09-252-991A-6953/c

; Sequence 6953, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6953

; LENGTH: 1848

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6953

Query Match

Best Local Similarity 3.1%; Score 41; DB 4; Length 1848;

Matches 100; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

QY 46 CGGGGGCGACCTCCGGGCGCGCGCGGAGCTGCAGCGCACTCTCGCGGCCAG 105

DB 1021 CTGGTGGCGCCCGCCACCGGCTCGGCCACCGCGGCTATCGGCAAGCTGCTGGCGGTGA 962

QY 106 GCGCGAGGTGTG-GAGAAAGCCCCSCACCTGGCAGCCCTAATTC-TGGCCCGGGAGGCGAG 164

DB 961 GGAACAGCGCGAGAGGCTCCAGGCTTGGCGCCACCGACCGCGCTGCGGGGTGA 902

QY 165 CAAGGCGATCCCTCTGAAGAACATTTAAGCACCTGGGGGGGTCCCGCTCATTTGGCTGGGT 224

DB 901 CGACCTGATGCTCGCTGGAGCGGCCACCGCTGGCGGGCTTGAGAGCCCCCAGGTCCGCGGGT 842

QY 225 C 225

DB 841 C 841

RESULT 13

US-09-252-991A-6781

; Sequence 6781, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6781

; LENGTH: 1899

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6781

Query Match	3.1%;	Score 41;	DB 4;	Length 1899;
Best Local Similarity	55.2%;	Pred. No. 0.071;		
Matches 100;	Conservative 0;	Mismatches 80;	Indels 1;	Gaps 1;
QY	46	CGGGGGCGACGTC	CGGGGGCGGGCGCCGCGCAAGCTC	CGGCAACTCTCGCGCGCGCCAG 105
Db	1017	CTGGTGGCCCGCA	CGGCGCTCGGCCACCGCGCATCGGC	AAGCTGCTGTGGCGGTGGA 1076
QY	106	GGCCGAGGTGTG	GGAGAGCCCGGCACCTGGCGAGACCCCTAA	TCTGGCCCGGGGAGGCGAG 164
Db	1077	GGAAAAAGAGCGCG	GAGGGTCTCCAGGCCCTGGGGGGCCACG	CGACGCGCCCTGCGGGGGTCTGA 1136
QY	165	CAAAGGCATCC	CGCTCGAAGAACTTAAGCACCTTGGGGGGGT	CCCGCTCATTTGGCTGGGT 224
Db	1137	CGACCTGATCC	CGCTTGAACGGCACCGTGGCGGGCTG	GGAGGCCCGCCAGGTCGCGGGGT 1196
QY	225	C	225	
Db	1197	C	1197	

RESULT 14

```

US-09-252-991A-13301/C
; Sequence 13301, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIORITY FILING DATE: 1993-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIORITY FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIORITY FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13301
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13301

```

Query Match	3.1%; Score 40.4; DB 4; Length 744;
Best Local Similarity	49.5%; Pred.No. 0.056;
Matches 104; Conservative 3; Mismatches 106; Indels 0; Gaps 0;	
QY	33 CGTCTCCAAACCGGGGCGGCACCGTCGCCGGGGCGGCGGAAGTGTCAGACGCAATC 92
DB	518 CTTTGGCCACAGTTTCGGGGTCCGAGCGCCGAGGTGCGTCTGTCGAACAGCATCATCTCCG 459
QY	93 TCGCGGGGCGAGGCGCCAGTGTGGAGAAGCCCCCGCACCTGGCAGCCCTTAATTCTGGC 152
DB	458 GCTCTCACCGCAGCGCGCGGCGCATGCCAGCGCTGCTGTCGCCCGCGCTCATGTGCG 399
QY	153 CCGGGAGGGCAGAAAGGCATCCCCTGAAGAACATTAAAGCACTGGCGGGGGCCCGCT 212
DB	398 CCGGGTAGGGCTCTTCCGGTGGCCACGCGGACTTGGCCAGGTAGTGTCTGGCCCTTCT 339
QY	213 CATTGGCTGGTCTCTCGTCGCGCCCTTGA 242
DB	338 CGATGGCTTCTTCTTCGACACGCCGAGCA 309

RESULT 15

US-09-252-991A-12981/c
Sequence 12981, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12981
; LENGTH: 783
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-12981

```

[illegible]

Search completed: May 11, 2004, 07:22:40
Job time : 129.426 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2004, 04:14:30 ; Search time 562.652 Seconds
(without alignments)
9853.151 Million cell updates/sec

Title: US-09-930-440B-3

Perfect score: 1305

Sequence: 1 atggactcgtggagaagg.....ataattcatgcacaaataag 1305

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001as.*

5: Geneseq2001bs.*

6: Geneseq2002s.*

7: Geneseq2003as.*

8: Geneseq2003bs.*

9: Geneseq2003cs.*

10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1303.4	99.9	1305	3	AAA50568 Human CMP
2	1303.4	99.9	1305	3	AAA37762 Human gly
3	1303.4	99.9	1305	4	AAB28457 Nucleotid
4	1303.4	99.9	1305	7	AA153992 DNA encod
5	1272.4	97.5	1770	4	AA158475 Human pol
6	1272.4	97.5	1770	8	ADB48450 Novel Hum
7	1272.4	97.5	1800	3	AA17722 Human can
8	1270.8	97.4	1694	4	AAH17722 Human cDN
9	1270.8	97.4	1751	3	AA177374 Human ORF
10	1270.8	97.4	1772	3	AA177374 Human car
11	1229.4	94.2	1803	4	AA160261 Human pol
12	586.8	52.6	738	8	ADB83231 Human cDN
13	547.6	49.6	789	8	ADB83156 Human cDN
14	636.6	48.8	838	6	ABQ89138 Human pro
15	627.4	48.1	745	8	ADB82082 Human cDN
16	625	47.9	819	6	ABQ88853 Human pro
17	620.2	47.5	748	6	ABQ89314 Human pro
18	620.2	47.5	748	8	ADB82257 Human cDN
19	617.4	47.3	708	8	ADB81797 Human cDN
20	562	43.1	686	6	ABQ89876 Human pro
21	562	43.1	686	8	ADB82915 Human cDN
22	555.6	42.6	639	6	ABQ98790 Human ORF
23	538.2	41.2	622	4	AAH06072 Human cDN

24	400.6	30.7	408	5	ACH49109	Ach49109 Human let
25	217	16.6	263	7	ABX44026	Abx44026 Bovine ES
26	181.6	13.9	589	4	AAH11143	Aah11143 Human cDN
27	145.8	11.2	222	6	ABN76702	Abn76702 Human ORF
28	145.4	11.1	205	7	ABX47106	Abx47106 Bovine ES
29	138.2	10.6	478	4	AA112541	Aa112541 Probe #24
30	138.2	10.6	478	4	ABA54246	Abas4246 Human fo
31	138.2	10.6	478	4	ABA43784	Abas43784 Human bre
32	138.2	10.6	478	4	ABA23995	Abas23995 Probe #24
33	138.2	10.6	478	4	AAK27959	Aak27959 Human bon
34	138.2	10.6	478	4	AAK02522	Aak02522 Human bra
35	138.2	10.6	478	4	ABS27544	Abas27544 Human liv
36	138.2	10.6	478	5	AA102446	Aa102446 Probe #24
37	134	10.3	134	4	AA121732	Aa121732 Probe #11
38	134	10.3	134	4	ABA66810	Abas66810 Human fo
39	134	10.3	134	4	ABA48889	Abas48889 Human bre
40	134	10.3	134	4	ABA33872	Abas33872 Probe #12
41	134	10.3	134	4	AAK40362	Aak40362 Human bon
42	134	10.3	134	4	AAK15239	Aak15239 Human bra
43	134	10.3	134	4	ABS40545	Abas40545 Human liv
44	134	10.3	134	5	AA107417	Aa107417 Probe #74
45	96.4	7.4	190	6	ABK35983	Abk35983 cDNA sequ

ALIGNMENTS

RESULT 1
AAA50568
ID AAA50568 standard; cDNA; 1305 BP.
XX
AC AAA50568;
XX
DT 19-DEC-2000 (first entry)
XX
DE Human CMP-sialic acid synthetase cDNA.
XX
KW CMP-sialic acid synthetase; human; sialylation; glycoprotein;
plasmalogen; transferrin; thyrotropin; Na+,K+-ATPase; ss.
XX
OS Homo sapiens.
XX
PN WO200052135-A2.
XX
PD 08-SEP-2000.
XX
PF 01-MAR-2000; 2000WO-US005313.
XX
PR 02-MAR-1999; 99US-0122582P.
PR 08-DEC-1999; 99US-0169624P.
XX
(EUNYA-) HUMAN GENOME SCI INC.
(UWJC) UNIV JOHNS HOPKINS.
(UWY-) UNIV WYOMING.
PI Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;
WPI; 2000-572178/53.
P-PSDB; AAY96100.
Recombinant production of sialylated glycoproteins using cells in which
the expression of enzymes, e.g. sialic acid synthetase, involved in the
sialylation reaction has been altered.
Disclosure; Page 100-102; 144pp; English.
The present sequence is that of human cDNA encoding CMP-sialic acid
synthetase (see AAY96100). This novel gene was identified on the basis of
homology with the Escherichia coli DNA sequence. The invention provides
methods and recombinantly engineered cells for producing glycoproteins
having sialylated oligosaccharides. The methods involve altering the
expression of enzymes involved in carbohydrate processing. A claimed cell
producing sialylated glycoprotein at above endogenous levels expresses at

XX PR 02-MAR-1999; 99US-0122409P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PT Coleman TA;
 XX DR WPI; 2000-572179/53.
 XX DR P-PSDB; AAY90351.
 XX PT New human glycosylation enzymes cytidine 5'-monophosphate sialic acid
 XX PT synthetase, sialic acid synthetase and aldolase and nucleic acids
 XX PT encoding the proteins for treating e.g., immune system disorders,
 XX PT microbial diseases.
 XX PS Claim 4; Page 106-108; 115pp; English.
 XX CC This sequence encodes a human glycosylation enzyme clone of the
 XX CC invention, designated HWLM34. The protein of this clone is a CMP sialic
 XX CC acid synthetase. The sequences are useful as reagents for the
 XX CC differential identification of the tissues or cell types present in a
 XX CC biological sample, as immunological probes, for treating a disease or
 XX CC condition resulting from under expression of such polypeptide, for the
 XX CC detection and/or treatment of disorders involving aberrant glycolysis,
 XX CC e.g. cramps, myoglobinuria, and as tumour marker and/or immunotherapy,
 XX CC targets. They may also be used to differentiate, proliferate and attract
 XX CC cells leading to the regeneration of tissues, to modulate mammalian
 XX CC characteristics (e.g. in cosmetic surgery) or mammalian metabolism
 XX CC affecting catabolism, anabolism processing, utilisation and energy
 XX CC storage, to change a mammal's mental state by influencing and as a food
 XX CC additive or preservative. The proteins can be used to assay protein
 XX CC levels in a sample, as a marker or detector of an immune system disorder,
 XX CC to inhibit cytokine activity, and as a vaccine. They may further be used
 XX CC to treat immune system or of haematopoietic cell deficiencies or
 XX CC disorders, blood coagulation disorders (e.g. afibrinogenemia), blood
 XX CC platelet disorders (e.g. thrombocytopenia), wounds resulting from trauma
 XX CC or surgery, autoimmune disorders (e.g. Addison's disease, multiple
 XX CC sclerosis, allergic encephalomyelitis), allergic reactions (e.g. asthma),
 XX CC organ rejection, graft-versus-host disease, inflammation,
 XX CC hyperproliferative disorders (e.g. neoplasia, purpura, sarcoidosis),
 XX CC diseases caused by viruses (e.g. hepatitis, meningitis, AIDS), bacteria
 XX CC and fungi (include e.g. tuberculosis, conjunctivitis, sepsis, typhoid,
 XX CC chlamydia, cellulitis), and diseases caused by parasites (e.g.
 XX CC amoebiasis, coccidiosis, leishmaniasis, scabies, malaria, toxoplasmosis)
 XX SQ Sequence 1305 BP; 382 A; 243 C; 350 G; 326 T; 0 U; 4 Other;
 Query Match 99.9%; Score 1303.4; DB 3; Length 1305;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGACTCGGTGGAGAGGGGGCGCCACCTCCGTCCTCCACCCCGGGGGCGACCGTCC 60
 DB 1 ATGACTCGGTGGAGAGGGGGCGCCACCTCCGTCCTCCACCCCGGGGGCGACCGTCC 60
 QY 61 CGGGGCGGGCGCGAAGCTCAGCGCAACTCTCGCGCGCGCGCGCGCGCGGGTGGAG 120
 DB 61 CGGGGCGGGCGCGAAGCTCAGCGCAACTCTCGCGCGCGCGCGCGCGGGTGGAG 120
 QY 121 AAGCCCGCGACCTCGGCGACCGCTAAATCTGGCCCGGGGAGGAGCAGCAAGGCGATCCCCCTG 180
 DB 121 AAGCCCGCGACCTCGGCGACCGCTAAATCTGGCCCGGGGAGGAGCAGCAAGGCGATCCCCCTG 180
 QY 181 AAGAACATTAAAGCACTGGCGGGGGTCCCGCTCATTTGGCTGGTCTCGGTGGCGCGCTG 240
 DB 181 AAGAACATTAAAGCACTGGCGGGGGTCCCGCTCATTTGGCTGGTCTCGGTGGCGCGCTG 240
 QY 241 GATTACAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCAGAAATG 300
 DB 241 GATTACAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCAGAAATG 300
 QY 301 GCCAAACAATTTGGTGACAGTTTCATCGAAGAAAGTTTCGAAGTTTCAAAGACAGCTCT 360

DB 301 GCCAAACAATTTGGTGCA CAAGTTTCATCGAAGAGTTCTGAAGTTTCAAAGACAGCTCT 360
 QY 361 ACCTCAGTAGATGCCATCATAGAAATTTCTTAATATATATATATATATATATATATATATAT 420
 DB 361 ACCTCAGTAGATGCCATCATAGAAATTTCTTAATATATATATATATATATATATATATATAT 420
 QY 421 AATATTCAGCTACTCTCTCATGTCTTACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
 DB 421 AATATTCAGCTACTCTCTCATGTCTTACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
 QY 481 ATTTCAGAGAGAGAGATATGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
 DB 481 ATTTCAGAGAGAGAGATATGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
 QY 541 AGTGAATTCAGAAAGAGGTTCTGTAAGTGCAGCACTCTGTAATTTAAATTCAGCTAA 600
 DB 541 AGTGAATTCAGAAAGAGGTTCTGTAAGTGCAGCACTCTGTAATTTAAATTCAGCTAA 600
 QY 601 CGGCTCTGTCGACAGACTGGGATGAGAAATTTATGAAATTTGCTTTTCTTTTCTTTTCTTTTCT 660
 DB 601 CGGCTCTGTCGACAGACTGGGATGAGAAATTTATGAAATTTGCTTTTCTTTTCTTTTCTTTTCT 660
 QY 661 AAAAGACATTTGATAGAGATGGTTTACTTGCAGGTGGAAATTTGCATATCTAGCAATTC 720
 DB 661 AAAAGACATTTGATAGAGATGGTTTACTTGCAGGTGGAAATTTGCATATCTAGCAATTC 720
 QY 721 GAGCTGGCACTAGTGTGGATATAGATGAGATTTGATTTGCTTTTCTTTTCTTTTCTTTTCTTTTCT 780
 DB 721 GAGCTGGCACTAGTGTGGATATAGATGAGATTTGATTTGCTTTTCTTTTCTTTTCTTTTCTTTTCT 780
 QY 781 GTATTAGATATGCTATTTTGGCAAGAGAGCTTAAAGAAATTTAAACTTTTGGTTTCT 840
 DB 781 GTATTAGATATGCTATTTTGGCAAGAGAGCTTAAAGAAATTTAAACTTTTGGTTTCT 840
 QY 841 AATATTGATGATGCTCAACATGCCACATTTATATATATATATATATATATATATATATATATAT 900
 DB 841 AATATTGATGATGCTCAACATGCCACATTTATATATATATATATATATATATATATATATATAT 900
 QY 901 ATATCTTATGATGATTAAGATGCTTATGGGATGATTTATTAAGAAAGTGGTATTTGAG 960
 DB 901 ATATCTTATGATGATTAAGATGCTTATGGGATGATTTATTAAGAAAGTGGTATTTGAG 960
 QY 961 GTAGGCTTAATCTCAGAAAGGCTTTCAGAGAGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
 DB 961 GTAGGCTTAATCTCAGAAAGGCTTTCAGAGAGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
 QY 1021 TGCATAATGGAAGTCAAGTGTATCAGCAAGCTAGCAGTTGTAGATGAATGGAGAAAGAA 1080
 DB 1021 TGCATAATGGAAGTCAAGTGTATCAGCAAGCTAGCAGTTGTAGATGAATGGAGAAAGAA 1080
 QY 1081 ATGGGCTTGTCTGAAAGAGAGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTGC 1140
 DB 1081 ATGGGCTTGTCTGAAAGAGAGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTGC 1140
 QY 1141 TTGAAGAGAGTGGCTTAAGTGGCGCT 1200
 DB 1141 TTGAAGAGAGTGGCTTAAGTGGCGCT 1200
 QY 1201 GTTGGATACATTTGCAATGTAATGGTGGCGCTTAAAGTGGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
 DB 1201 GTTGGATACATTTGCAATGTAATGGTGGCGCTTAAAGTGGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
 QY 1261 ATTTGCTCTTAATTCGAAAGAGTGAATTAATTCATGCCAAATATAG 1305
 DB 1261 ATTTGCTCTTAATTCGAAAGAGTGAATTAATTCATGCCAAATATAG 1305
 RESULT 3
 AAH28457
 ID AAH28457 standard; DNA; 1305 BP.
 XX AC
 XX AH28457;
 XX

D- 17-SRP-2001 (first entry)
 XX Nucleotide sequence of a cytidine monophosphate-sialic acid synthetase.
 DE
 XX Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;
 KW cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;
 KM sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;
 KW vaccine; ss.
 XX Homo sapiens.
 OS
 XX
 FH Key
 FT CDS
 FT 1..1305
 FT /tag= a
 FT /transl_except= (pos: 397..399, aa: Xaa)
 FT /transl_except= (pos: 406..408, aa: Xaa)
 FT /transl_except= (pos: 439..441, aa: Xaa)
 FT /transl_except= (pos: 505..507, aa: Xaa)
 FT /product= "cytidine monophosphate-sialic acid synthetase"
 FT /note= "Xaa is an unspecified residue"
 XX
 XX WO200142492-A1.
 XX
 XX 14-JUN-2001.
 XX
 XX 07-DEC-2000; 2000MC-US033136.
 XX
 XX 09-DEC-1999; 99US-0169839P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX (UYJO) UNIV JOHNS HOPKINS.
 XX (UTEM) UNIV TEMPLE.
 XX (UYWV-) UNIV WYOMING.
 XX
 XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Falter K;
 PI Jarvis D;
 PI WPI: 2001-441575/47.
 DR P-PSDB; AAB84682.
 DR
 XX
 XX Cells producing cytidine monophosphate-sialic acid and sialylated
 FT glycoprotein above endogenous levels for production of vaccines and
 FT therapeutics.
 FT
 XX Disclosure; Fig 29; 182pp; English.
 PS
 XX The specification describes a method for manipulating carbohydrate
 CC processing pathways in cells of interest. The methods are used to
 CC manipulate multiple pathways involved with the sialylation reaction by
 CC using recombinant DNA technology and substrate feeding approaches to
 CC enable the production of sialylated glycoproteins in the cells. The
 CC sialylation process involves the post-translational addition of the donor
 CC substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific
 CC acceptor carbohydrate. The cells express at least one enzyme, selected
 CC from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,
 CC CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The
 CC cells are useful for producing complex sialylated glycoproteins in cells
 CC of interest, especially insect cells. Glycoproteins containing sialylated
 CC oligosaccharides are useful as vaccines, therapeutics and diagnostic
 CC tools. Cells producing complex sialylated glycoproteins are useful for
 CC enhancing the value of heterologous expression systems and increasing the
 CC application of heterologous cell expression products as vaccines,
 CC therapeutics and diagnostic tools as well as increasing the variety of
 CC heterologous proteins that can be produced and lowering biotechnology
 CC production costs. The present sequence encodes a human CMP-SA synthetase,
 CC which is used in the method of the invention
 XX
 XX Sequence 1305 BP; 382 A; 243 C; 350 G; 326 T; 0 U; 4 Other;
 SQ
 Query Match 99.9%; Score 1303.4; DB 4; Length 1305;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

	D _b	341	GCCAAACAATTGGTGCACAGTTTCATCBAGAAGTCTGGAAGTTTCAAAGACAGCTCT	400
	Q _y	361	ACCTCAGTAGATGCCATCATAGAAATTTCTTAATTATYATATGAGSGKTGCANTGTAGGA	420
	D _b	401	ACCTCAGTAGATGCCATCATAGAAATTTCTTAATTATCATATGAGGTTGACATTGTAGGA	460
	Q _y	421	AATATTCAAGCTACTCTTCYTCHATTGTACACTCTACTGATCTTCAAAAAGTTGCAGAAATG	480
	D _b	461	AATATTCAAGCTACTCTTCCAATGTTTATCATCTCTACTGATCTTCAAAAAGTTGCAGAAATG	520
	Q _y	481	ATTGCGAAGAAGGATATGATTCTGKTTTTCTCTGTGTGACAGCCCATCAGTTTCGATGG	540
	D _b	521	ATTGCGAAGNAGGATATGATTCGTITTTCTGTGTGTGACAGCCCATCAGTTTCGATGG	580
	Q _y	541	AGTGAAATTCAGAAAGGAGTTTCGTGAAGTGCACCGAACCTCTGMAATTTAMATCCAGCTAAA	600
	D _b	581	AGTGAAATTCAGAAAGGAGTTCGTGAAGTGCACCGAACCTCTGMAATTTAMATCCAGCTAAA	640
	Q _y	601	CGGCTCGTGCACAAAGACTGGGATGCGAATATATAGNAATGGCTCANTTATTTGCT	660
	D _b	641	CGGCTCGTGCACAAAGACTGGGATGCGAATATATAGNAATGGCTCANTTATTTGCT	700
	Q _y	661	AAAGACATTTGATAGAGATGGGTACTTTCAGAGGGTGG-AAAAATGGCATACTAGCAAAATG	719
	D _b	701	AAAAGACATTTGATAGAGATGGGTACTTTCAGAGGGTGGAAAAATGGCATACTATGAAATG	760
	Q _y	720	CGAGCTGSHACATAGTGTGNATPAGATGTGGATATTGATTTGGCTTATTCGACAGACAAAG	779
	D _b	761	CGAGCT--GNAACATAGTGTGGATAPAGATGTGGATATTGATTTGGCTTATTCGACAGACAAAG	819
	Q _y	780	AETATTAAAGATATGGCTATTTTTGGCAAGAGAAGCTTAAGGAAATAAAACTTTTGGTTTG	839
	D _b	820	AGTATTAAAGATATGGCTATTTTTGGCAAGAGAAGCTTAAGGAAATAAACTTTTGGTTTG	879
	Q _y	840	CATATTGATGGATGTCTCACCAATGGCCACATTTATGTATCATGAGACACAAAAGAAAT	899
	D _b	880	CATATTGATGGATGTCTCACCAATGGCCACATTTATGTATCATGAGACACAAAAGAAAT	939
	Q _y	900	PACATCTTATGATGTAAAGATGCTATTGGGATAAGTTTATTAAGAAAAAGTGTATTGA	959
	D _b	940	AHTATCTTATGATGTAAAGATGCTATTGGGATAAGTTTATTAAGAAAAAGTGTATTGA	999
	Q _y	960	GGTGAGGCTAATCTCAGAAAAGGSCCTGTTCAAAGCAGACGCTGTCTCTTTTAAAACTGGA	1019
	D _b	1000	GGTGAGGCTAATCTCAGAAAAGGSCCTGTTCAAAGCAGACGCTGTCTCTTTTAAAACTGGA	1059
	Q _y	1020	TTCABAATCGNAGTCACTGTATCAGACACAGCTGACGTTGTAGATGANTGGAGAAAGA	1079
	D _b	1060	TTCABAATCGNAGTCACTGTATCAGACACAGCTGACGTTGTAGATGANTGGAGAAAGA	1119
	Q _y	1080	AATGGGCTGTGCTGGAAGAAGTGGCATATCTTGGAAATGAGAGTCTCATGAGAGAGTG	1139
	D _b	1120	AATGGGCTGTGCTGGAAGAAGTGGCATATCTTGGAAATGAGAGTCTCATGAGAGAGTG	1179
	Q _y	1140	CTTGAGAGAGTGGGCTTAAGTGCGBCTCCHGTGATGCTCTGTCTACCGCCCGAAGGC	1199
	D _b	1180	CTTGAGAGAGTGGGCTTAAGTGCGBCTCCTGTGATGCTCTGTCTACTGCCCAGAGGC	1239
	Q _y	1200	TGTTGGATACATTTTGCAAAATGTAATGTTGGCCGTGGTGGCAATCCGAGAATTTGCAGAGCA	1259
	D _b	1240	TGTTGGATACATTTTGCAAAATGTAATGTTGGCCGTGGTGGCAATCCGAGAATTTGCAGAGCA	1299
	Q _y	1260	CATTGCTCTACTANTGGAAGAAGTTAATAATCATATGCCAAAATAG	1305
	D _b	1300	CATTGCTCTACTANTGGAAGAAGTTAATAATCATATGCCAAAATAG	1345

RESULT 9

AAC77374
ID AAC77374 standard; cDNA; 1751 BP.

AAC77374:

08-FEB-2001 (first entry)

Human ORFX ORF2329 polynucleotide sequence SEQ ID NO:5857.

Human; open reading frame; ORF; detection; cytostatic; hepatotropic; vulnery; antiparasitic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antihypertoid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000: 200000-US0008621-

31-MAP-1999. 99ITS-0127607P

02-APR-1999; 99US-0127636P.
05-APR-1999; 99US-0127728P.

30-MAR-2000; 2000US-00540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M

WPI; 2000-602362/57.

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

useful for treati

Claim 5; Page 5027-5028; 5507pp; English.

AAC74446 to AAC77606 encode the proteins

which represent the human ORF open reading frames 1 to 304. The sequences have activities such as: cytostatic; hepatotropic; vulnery; antispasitic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cadiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antineumatic; antitumor; and antinaeamic. The sequences can be used for determining the presence of or predisposition to, ORF-associated disorder. The nucleic acids can be used to express ORF proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive

97.44: Score 1270.8: DB 3: Length 1751;

Very Match	37.44	Score	1270.8
1st Local Similarity	39.14	Pred. No.	0

st local similarity 95-100; Gaps 2; Indels 6; Mismatches 4; Conservative matches 1294; Mismatches 4; Indels 6; Gaps 2;

RESULT 9

AAC77374
ID AAC77374 standard; cDNA; 1751 BP.

AAC77374:

QY	1	ATGGACTGGTGGAGAGGGGGCCGACCTCGGCTCCAAACCCGCGGGGGCGACCGTCC	60
Db	76	ATGGACTGGTGGAGAGGGGGCCGACCTCGGCTCCAAACCCGCGGGGGCGACCGTCC	135
QY	61	CGGGCCGGCCCGCCGAGCTGCAGCGCACTCTCGGGGGGCGCAGGGCGAGGTGGNG	120
Db	136	CGGGCCGGCCCGCCGAGCTGCAGCGCACTCTCGGGGGGCGCAGGGCGAGGTGGNG	195
QY	121	AGCCGCCCGCAGCTGGCGACCCCTAATTTCTGGCCCGGGAGGCGAGCAAAAGGCATCCCCCTG	180
Db	196	AGCCGCCCGCAGCTGGCGACCCCTAATTTCTGGCCCGGGAGGCGAGCAAAAGGCATCCCCCTG	255
QY	181	AAGAACTATAGCACTGCGCGGGGTCCCGCTCATTTGGCTGGGTCTGGTGGCGGCCCTG	240
Db	256	AAGAACTATAGCACTGCGCGGGGTCCCGCTCATTTGGCTGGGTCTGGTGGCGGCCCTG	315
QY	241	GAATCAGGGGCTTCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCAGAGATGG	300
Db	316	GAATCAGGGGCTTCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCAGAGATGG	375
QY	301	GCCAAACAATTTGGTGCACAGTTTCATCGAAGAAGTTCTGAAGTTTCAAAAGACAGCTCT	360
Db	376	GCCAAACAATTTGGTGCACAGTTTCATCGAAGAAGTTCTGAAGTTTCAAAAGACAGCTCT	435
QY	361	ACCTCACTAGATGCCATCATAGAAATTTCTTAATTATATATTAATGAGGKTGACATTGTAGA	420
Db	436	ACCTCACTAGATGCCATCATAGAAATTTCTTAATTATATATTAATGAGGKTGACATTGTAGA	495
QY	421	AATATTCAGCTACTCTTCATGTTTACATCCTACTGATCTTCAAAAAGTTCCAGAAATG	480
Db	496	AATATTCAGCTACTCTTCATGTTTACATCCTACTGATCTTCAAAAAGTTCCAGAAATG	555
QY	481	ACTCGAGAAGAGGATATGATTCGTTCTCTGTTGTGAGCGCCATCAGTTTCGATGG	540
Db	556	ACTCGAGAAGAGGATATGATTCGTTCTCTGTTGTGAGCGCCATCAGTTTCGATGG	615
QY	541	AGTGAATTCAGAAAGGATTCGTGAGTGCAGGACCGAACTCTGAATTTAAATCCAGCTAA	600
Db	616	AGTGAATTCAGAAAGGATTCGTGAGTGCAGGACCGAACTCTGAATTTAAATCCAGCTAA	675
QY	601	CGGCTCGTCCAGCAAGCTGGGATCGAGAAATATATGAAATGGCTCATTTTATTTGCT	660
Db	676	CGGCTCGTCCAGCAAGCTGGGATCGAGAAATATATGAAATGGCTCATTTTATTTGCT	735
QY	661	AAAAGACATTTGATAGAGATGGTTTACTTGAGGCTGG-AAAATGGCATCTACGAAATG	719
Db	736	AAAAGACATTTGATAGAGATGGTTTACTTGAGGCTGGAAATATGCAATCTACGAAATG	795
QY	720	CGAGCTGAACATAGTGTGATATAGATGTGATATTCATTTGGCTTATTCAGAGCAAG	779
Db	796	CGAGCT-GAATAGTGTGATATAGATGTGATATTCATTTGGCTTATTCAGAGCAAG	854
QY	780	AGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAGGAAATAAACCTTTTGGTTG	839
Db	855	AGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAGGAAATAAACCTTTTGGTTG	914
QY	840	CAATATTGATGATGTCACCAATGGCCACATTTATGTTATCAGGAGCCAAAAGAAAT	899
Db	915	CAATATTGATGATGTCACCAATGGCCACATTTATGTTATCAGGAGCCAAAAGAAAT	974
QY	900	AATATCTTATGATGTAAAGATGCTATTGGGATAGTTTATTAAGAAAGTGGTATTGA	959
Db	975	AATATCTTATGATGTAAAGATGCTATTGGGATAGTTTATTAAGAAAGTGGTATTGA	1034
QY	960	GGTGAGGCTAATCTCAGAAAGGCGCTGTTCAAAGCAGACGGTGTCTTCTTTAAACTGGA	1019
Db	1035	GGTGAGGCTAATCTCAGAAAGGCGCTGTTCAAAGCAGACGGTGTCTTCTTTAAACTGGA	1094
QY	1020	TTGCATAATGGAAGTCACTGTATCAGCAAGCTAGCTGTAGATGAATGGAAGAAAGA	1079
Db	1095	TTGCATAATGGAAGTCACTGTATCAGCAAGCTAGCTGTAGATGAATGGAAGAAAGA	1154
QY	1080	AATGGGCTGTGCTGGAAAGAGTGGCCATCTCTTGGAAATGAAGTCTCTGATGAAGATG	1139
Db	1155	AATGGGCTGTGCTGGAAAGAGTGGCCATCTCTTGGAAATGAAGTCTCTGATGAAGATG	1214
QY	1140	CTTGAAGACAGTGGGCTTAAGTGGCGCTCCTGCTGATGCTGTTCCTAGCCAGAGAGGC	1199
Db	1215	CTTGAAGACAGTGGGCTTAAGTGGCGCTCCTGCTGATGCTGTTCCTAGCCAGAGAGGC	1274
QY	1200	TGTTGATACATTTGCAAAATGTAATGGTGGCGCTGCTGCTGATGCTGCTGCTGCTGCTG	1259
Db	1275	TGTTGATACATTTGCAAAATGTAATGGTGGCGCTGCTGCTGATGCTGCTGCTGCTGCTG	1334
QY	1260	CAATTTGCTTACTTAATGGGAAAGATTAATTAATTCATGCTCCAAATATAG	1305
Db	1335	CAATTTGCTTACTTAATGGGAAAGATTAATTAATTCATGCTCCAAATATAG	1380
RESULT 10			
AAC65392			
ID	AAC65392 standard; cDNA; 1772 bp.		
XX	AAC65392;		
AC	13-FEB-2001 (first entry)		
DT	Human carbohydrate-modifying enzyme cDNA Incyte ID No: 000422CB1.		
XX	Human; carbohydrate-modifying enzyme; CME; antidiabetic;		
KW	immunosuppressive; anti-HIV; anti-inflammatory; antianemic;		
KW	antiasthmatic; antiarteriosclerotic; antithyroid; hepatotropic;		
KW	nephrotropic; antigout; thymimetic; neuroprotective; osteopathic;		
KW	antiarthritic; antipruritic; uropathic; ophthalmological;		
KW	dermatological; antitumor; cytostatic; virucide; antibacterial;		
KW	fungicide; protozoicide; tranquilizer; vulnerary; diabetes;		
KW	autoimmune disorder; inflammatory disorder; infection; ss.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	WO200063351-A2.		
FN	26-OCT-2000.		
XX	20-APR-2000; 2000WO-US010882.		
PF	21-APR-1999; 99US-0130383P.		
XX	{INCY-} INCYTE GENOMICS INC.		
PA	Lal P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J;		
XX	WPI; 2000-672729/65.		
DR	P-PSDB; AAB28673.		
XX	Novel carbohydrate modifying enzyme polypeptides and polynucleotides for		
PT	diagnosis, treatment, and prevention of carbohydrate metabolism		
PT	disorders, autoimmune/inflammatory disorders, and cancer.		
XX	Claim 4; Page 72; 75pp; English.		
PS	The present cDNA sequence encodes a human carbohydrate-modifying enzyme		
XX	(CME). CME polynucleotides and polypeptides are useful for treating and		
CC	diagnosing diseases associated with CME such as diabetes, adult		
CC	autoimmune/inflammatory disorders such as AIDS, Addison's disease, adult		
CC	respiratory distress syndrome, allergies, anaemia, asthma,		
CC	atherosclerosis, autoimmune thyroiditis, bronchitis, cholecystitis,		
CC	contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,		
CC	glomerulonephritis, multiple sclerosis, myasthenia gravis,		
CC	Hashimoto's thyroiditis, Good pasture's syndrome, gout, Grave's disease,		
CC	osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,		
CC	Reiter's syndrome, arthritis, scleroderma, Sjogren's syndrome, systemic		
CC	lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,		
CC	complications of cancer, haemodialysis, and extracorporeal circulation,		
CC	viral, bacterial, fungal parasitic, protozoal, and helminthic infections,		
CC	trauma, or cancer. CME, or its catalytic or immunogenic fragment, is		

CC useful for drug screening

XX Sequence 1772 BP; 516 A; 311 C; 464 G; 481 T; 0 U; 0 Other;

Query Match 97.4%; Score 1270.8; DB 3; Length 1772;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1294; Conservative 4; Mismatches 6; Indels 2; Gaps 2;

QY 1 ATGGAGCTGGTGGAGAGGGGCGCCACCTCGCTCCAAACCGGGGGGAGCGCTCC 60
DB 101 ATGGAGCTGGTGGAGAGGGGCGCCACCTCGCTCCAAACCGGGGGGAGCGCTCC 160
QY 61 CGGGGCGGGCGCCGAGCTGCGAGCGCACTCTGCGGGCGGCGAGCGGCGAGCTGTGGAG 120
DB 161 CGGGGCGGGCGCCGAGCTGCGAGCGCACTCTGCGGGCGGCGAGCGGCGAGCTGTGGAG 220
QY 121 AAGCCCCCGACCTGCGAGCCCTAATCTTGGCCCGGGGAGGCGAGAAAGGCAATCCCCCTG 180
DB 221 AAGCCCCCGACCTGCGAGCCCTAATCTTGGCCCGGGGAGGCGAGAAAGGCAATCCCCCTG 280
QY 181 AAGACATTAAGCACTGCGGGGGTCCCGCTCATTTGGCTGGTCCGCTGGCGGCGCTG 240
DB 281 AAGACATTAAGCACTGCGGGGGTCCCGCTCATTTGGCTGGTCCGCTGGCGGCGCTG 340
QY 241 GATTCAGGGGCTTCCAGAGTGTATGGGTTTCCAGAGCCATCATGAAATTCAGAAATG 300
DB 341 GATTCAGGGGCTTCCAGAGTGTATGGGTTTCCAGAGCCATCATGAAATTCAGAAATG 400
QY 301 GCGAAACATTTGGTGCACAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTT 360
DB 401 GCGAAACATTTGGTGCACAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTT 460
QY 361 ACCTCAGTAGATGCGCATAGAAATTTCTTAATATATATATATATATATATATATATAT 420
DB 461 ACCTCAGTAGATGCGCATAGAAATTTCTTAATATATATATATATATATATATATATAT 520
QY 421 AATATTCAGTACTTCTCATGTTTACATCTTCTGATCTTCAAAAGTTTGCAGAAATG 480
DB 521 AATATTCAGTACTTCTCATGTTTACATCTTCTGATCTTCAAAAGTTTGCAGAAATG 580
QY 481 ATTCCAGAAAGAGATATGATTTCTGTTCTCTGTTGTGAGAGCGCATCAGTTTCTGATGG 540
DB 581 ATTCCAGAAAGAGATATGATTTCTGTTCTCTGTTGTGAGAGCGCATCAGTTTCTGATGG 640
QY 541 AGTGAATTCAGAAAGAGTCTGTAAGTGAAGGAGCGACCTCTGAATTTAATCCAGCTAA 600
DB 641 AGTGAATTCAGAAAGAGTCTGTAAGTGAAGGAGCGACCTCTGAATTTAATCCAGCTAA 700
QY 601 CGGCTCTGTCGACAGACTGGATGAGAAATATATGAAATGGCTCATTTTATTTGCT 660
DB 701 CGGCTCTGTCGACAGACTGGATGAGAAATATATGAAATGGCTCATTTTATTTGCT 760
QY 661 AAAGACATTTGATAGAGTGGTTACTTTCAGGGTGG-AAAATGSCATCTACGAAATG 719
DB 761 AAAGACATTTGATAGAGTGGTTACTTTCAGGGTGGAAATGCGCATCTATGAAATG 820
QY 720 CGAGCTGGAACATAGCTGGATATAGATGTGATATGATGCGCTATTGCGAGAGCAAG 779
DB 821 CGAGCT-GAACATAGCTGGATATAGATGTGATATGATGCGCTATTGCGAGAGCAAG 879
QY 780 AGATTAAGATAGCTATTTTGGCAAGAGAGCTTAAGGAAATATAACTTTTGGTTTG 839
DB 880 AGATTAAGATAGCTATTTTGGCAAGAGAGCTTAAGGAAATATAACTTTTGGTTTG 939
QY 840 CACATTTGATGATGCTCTACCAATGCGGACATTTATGATCATGAGAGCAAAAGAAAT 899
DB 940 CACATTTGATGATGCTCTACCAATGCGGACATTTATGATCATGAGAGCAAAAGAAAT 999
QY 900 AATATCTTATGATTAAGAGATGCTATTGGGATAAGTTTATTAAGAAAGTGGATTCA 959
DB 1000 AATATCTTATGATTAAGAGATGCTATTGGGATAAGTTTATTAAGAAAGTGGATTCA 1059
QY 960 GGTGAGGCTAATCTCAGAAAGGGGCTGTTCAAAGCAGACGCTGTCTCTTTAAACCTGGA 1019

DB 1060 GGTGAGGCTAATCTCAGAAAGGGGCTGTTCAAAGCAGACGCTGTCTCTTTAAACCTGGA 1113
QY 1020 TTGCAAAATGGAAGTCAGTGTATTCAGACAAAGCTAGCAGTTGTAGTGAATGGGAAAAGA 1079
DB 1120 TTGCAAAATGGAAGTCAGTGTATTCAGACAAAGCTAGCAGTTGTAGTGAATGGGAAAAGA 1179
QY 1080 AATGGGCTGTCTCGAAGAGAGTGGCTATCTTGAATGAAGTGTCTGATGAAGAGTG 1139
DB 1180 AATGGGCTGTCTCGAAGAGAGTGGCTATCTTGAATGAAGTGTCTGATGAAGAGTG 1239
QY 1140 CTTTGAAGAGAGTGGGCTTAAGTGGGCTCTCTGCTGATGCTGTCTTCTACGCCAGAGGC 1199
DB 1240 CTTTGAAGAGAGTGGGCTTAAGTGGGCTCTCTGCTGATGCTGTCTTCTACGCCAGAGGC 1299
QY 1200 TGTGATGATCATTTGCAAAATGTAATGCTGGCGCTGTGTCATCCGAGAAATTCGAGAGCA 1259
DB 1300 TGTGATGATCATTTGCAAAATGTAATGCTGGCGCTGTGTCATCCGAGAAATTCGAGAGCA 1359
QY 1260 CATTTGCTTACTTAATGGAAGAGTTAATTAATTCATGCCAAATAAG 1305
DB 1360 CATTTGCTTACTTAATGGAAGAGTTAATTAATTCATGCCAAATAAG 1405

RESULT 11
AAI60261/c
ID AAI60261 standard; cDNA; 1803 BP.
XX
AC AAI60261;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4250.
XX
DE Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang Z, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QX;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41105.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Claim 1; SEQ ID NO 4250; 10078pp; English.
XX

CC The invention relates to human nucleic acids (AA157798-AA161369) and the
 CC encoded polypeptides (AA138642-AA142213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX
 SQ Sequence 1803 BP; 483 A; 461 C; 330 G; 529 T; 0 U; 0 Other;

Query Match 94.2%; Score 1229.4; DS 4; Length 1803;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1274; Conservative 4; Mismatches 5; Indels 4; Gaps 4;

QY 1 ATGACTCGTGGAGAGGGGCGCCACCTCGCTCCCAACCCGCGGGGCGACCGTCC 60
 Db 1699 ATGACTCGTGGAGAGGGGCGCCACCTCGCTCCCAACCCGCGGGGCGACCGTCC 1640

QY 61 CGGSCCGGCGCCGAGAGCTGAGCGCAACTCTCGCGGGCGGCGAGGCGAGGTGGAG 120
 Db 1639 CGGSCCGGCGCCGAGAGCTGAGCGCAACTCTCGCGGGCGGCGAGGCGAGGTGGAG 1580

QY 121 AAGCCCGCGACCTGGAGCCCTAATCTCGCGCGGGGAGGAGCAAGGCAATCCGCCG 180
 Db 1579 AAGCCCGCGACCTGGAGCCCTAATCTCGCGCGGGGAGGAGCAAGGCAATCCGCCG 1520

QY 181 AAGAACATTAAACACTCGCGGGGCTCGCGCTCATTTGGTGGGTCTCGTGGCGGCCG 240
 Db 1519 AAGAACATTAAACACTCGCGGGGCTCGCGCTCATTTGGTGGGTCTCGTGGCGGCCG 1460

QY 241 GATTGAGGGGCTCCAGAGTGTATGGGTTTCGACAGACCATGTA-TAAATTGAGAAATG 300
 Db 1459 GATTGAGGGGCTCCAGAGTGTATGGGTTTCGACAGACCATGTA-TAAATTGAGAAATG 1400

QY 301 GCGAAGCAATTGGTGACAGATGTCATCGAAGAGTCTGAGAGTTCGAAAGACAGCTCT 360
 Db 1399 GCGAAGCAATTGGTGACAGATGTCATCGAAGAGTCTGAGAGTTCGAAAGACAGCTCT 1340

QY 361 ACCTCACTAGATGCATCATAGAAATTTCTTAATTATATATGAGGKTGACATCTAGGA 420
 Db 1339 ACCTCACTAGATGCATCATAGAAATTTCTTAATTATATATGAGGKTGACATCTAGGA 1280

QY 421 AATATCAAGCTACTCTTCATGTTTATCATCTACTGATCTTCAAAAAGTTCAGAAATG 480
 Db 1279 AATATCAAGCTACTCTTCATGTTTATCATCTACTGATCTTCAAAAAGTTCAGAAATG 1220

QY 481 ATTGAGAGAGGATATGATCTGKTTCTCTGTTGAGAGCCATCAGTTTCGATGG 540
 Db 1219 ATTGAGAGAGGATATGATCTGKTTCTCTGTTGAGAGCCATCAGTTTCGATGG 1160

QY 541 AGTCAAAATTGAGAAAGGATTCGTTGAAGTGCAGCAACCTCTGAATTTAAATCCAGCTAAA 600
 Db 1159 AGTCAAAATTGAGAAAGGATTCGTTGAAGTGCAGCAACCTCTGAATTTAAATCCAGCTAAA 1100

QY 601 CGGCTCGTCAAGACTGGGATGAGAAATPATATGAAATGGCTCATTTTATTTGCT 660
 Db 1099 CGGCTCGTCAAGACTGGGATGAGAAATPATATGAAATGGCTCATTTTATTTGCT 1040

QY 661 AAAGACATTTGATAGATGGTTACTTTCAGAGGTGG-AAATGGCACTACAGAAATG 719
 Db 1039 AAAGACATTTGATAGATGGTTACTTTCAGAGGTGGAAATGGCACTACAGAAATG 980

QY 720 CGAGCTGGAACATAGTGGGATACAGATGTGATATTGATTGGGCTATTGAGAGCAAG 779
 Db 379 CGAGCT-GAACATAGTGGGATACAGATGTGATATTGATTGGGCTATTGAGAGCAAG 921

QY 780 AGTATTAGATATGGCTATTGTTGGCAAGAGAGCTTAAGGAATAAATCTTTGTTG 839
 Db 920 AGTATTAGATATGGCTATTGTTGGCAAGAGAGCTTAAGGAATAAATCTTTGTTG 861

QY 840 CAATATTGATGATGCTCCCAATGGCCGCACTTTATGTTATCGAGAGACCAAAAGAAAT 899
 Db 860 CAATATTGATGATGCTCCCAATGGCCGCACTTTATGTTATCGAGAGACCAAAAGAAAT 801

QY 900 AATATCTTATGATGATGTTAAAGAGTCTATTGGGTAAAGTATTAAACAAAGTCTATTGA 959
 Db 800 AATATCTTATGATGATGTTAAAGAGTCTATTGGGTAAAGTATTAAACAAAGTCTATTGA 741

QY 960 GGTGAGGCTTAATCTCAGAAAGGCGCTGTTTCAAGCAGACGCTGTTCTTTTAAACTGGA 1019
 Db 740 GGTGAGGCTTAATCTCAGAAAGGCGCTGTTTCAAGCAGACGCTGTTCTTTTAAACTGGA 681

QY 1020 TTGCAAAATGGAAGTCAGTGTATCAGACAAAGCTAGAGTTGTAGTGAATGGAAGAAGA 1079
 Db 680 TTGCAAAATGGAAGTCAGTGTATCAGACAAAGCTAGAGTTGTAGTGAATGGAAGAAGA 621

QY 1080 AATGGGCTGCTGCTGGAAGAGAGTGGCATATCTTGGAAATGAATCTCTGTATGCAAGAGTG 1139
 Db 620 AATGGGCTGCTGCTGGAAGAGAGTGGCATATCTTGGAAATGAATCTCTGTATGCAAGAGTG 561

QY 1140 CTTGAGAGAGTGGGCTTAAGTGGGCTCCTGCTGATGCTGTTCTTACGCCCAAGAGGC 1199
 Db 560 CTTGAGAGAGTGGGCTTAAGTGGGCTCCTGCTGATGCTGTTCTTACGCCCAAGAGGC 501

QY 1200 TGTGATGATATCTTCAAAATGTAATGTTGGTGGCGTGGTG-CCATCCAGAAATTCAGAGC 1258
 Db 500 TGTGATGATATCTTCAAAATGTAATGTTGGTGGCGTGGTG-CCATCCAGAAATTCAGAGC 441

QY 1259 ACATTG-CCTACTAATGGAAGAT 1284
 Db 440 ACATTG-CCTACTAATGGAAGAT 414

RESULT 12
 ADB83231

ID ADB83231 standard; cDNA; 738 BP.

XX ADB83231;

XX 04-DEC-2003 (first entry)

XX Human cDNA sequence useful for the treatment of cancer (SeqID 1444).

XX human; prostate; cancer; cytostatic; gene therapy; vaccine;

XX immune response; gene; ss.

XX Homo sapiens.

XX WO2003050236-A2.

XX 19-JUN-2003.

XX 04-SEP-2002; 2002WO-US028214.

XX 07-DEC-2001; 2001US-00012697.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;

PI Crkvenjakov R, Dickson M, Drmanac S, Labat I, Leshkowitz D, Kita D;

PI Garcia V, Jones LM, Stache-Crain B, Scott EM;

PI WPI; 2003-513972/48.

XX New polynucleotides derived from human prostate, useful for modulating
 XX immune response to prevent or treat cancer.

PS	Claim 1; SEQ ID NO 1444; 188pp; English.	
XX	This invention relates to novel isolated polynucleotides of human origin, particularly isolated from the human prostate. Specifically, it refers to the diagnostics and therapeutics comprising these novel human polynucleotides, and includes the derived probes, antisense oligonucleotides and antibodies thereof. The identification of these human prostate genes that can inhibit tumour growth is useful for understanding the progression and nature of complex diseases such as cancer, and hence they are important in the drug discovery process. The present invention describes these polynucleotides and encoded polypeptides as exhibiting cytostatic activity, and through gene therapy and/or vaccines they can be used to modulate the immune response for the prevention or treatment of cancers, particularly of the prostate, but also for breast, lung and colon cancer. This polynucleotide sequence is a human cDNA sequence useful for the treatment of cancer, used in an exemplification of the invention. NOTE: These sequences are not given in the specification but are provided on the WIPO website.	
XX	Sequence 738 BP; 247 A; 97 C; 179 G; 215 T; 0 U; 0 Other;	
SQ	Query Match 52.6%; Score 686.8; DB 8; Length 738; Best Local Similarity 97.3%; Pred. No. 3.2e-184; Matches 716; Conservative 3; Mismatches 15; Indels 2; Gaps 2;	
QY	405 GGGTGCATTTGAGGAAATATTCAGCTACTCTTCYCATGTTTACATCTCTGATCTTCA 464	
DB	3 GGTGACATTTAGGAATATTCAGCTACTCTCCCATGTTTACATCTCTGATCTTCA 62	
QY	465 AAAAGTTCGAAATGATTCGAGNAGAGATATGATCTGKTTCTCTGTGTGAGAG 524	
DB	63 AAAAGTTCGAAATGATTCGAGNAGAGATATGATCTGKTTCTCTGTGTGAGAG 122	
QY	525 CCATCAGTTTCGATGGAGTGAATTCAGAAAGAGTTTCGTAAGTGCAGCAACCTTCAA 584	
DB	123 CCATCAGTTTCGATGGAGTGAATTCAGAAAGAGTTTCGTAAGTGCAGCAACCTTCAA 182	
QY	585 TTTAAATCCAGTAAACGGCTCTGACAGAGATCGGATGGAGATATATGAAATGG 644	
DB	183 TTTAAATCCAGTAAACGGCTCTGACAGAGATCGGATGGAGATATATGAAATGG 242	
QY	645 CTGATTTATTTGCTAAAAGACATTTGATAGATGGTTACTTCAGGGTGG-AAAAT 703	
DB	243 CTGATTTATTTGCTAAAAGACATTTGATAGATGGTTACTTCAGGGTGG-AAAAT 302	
QY	704 GGCACTACGAAATCGAGCTGCAACATATGTTGGATATAGATGTGATATTGATTGSC 763	
DB	303 GGCACTACGAAATCGAGCT-GAACATATGTTGGATATAGATGTGATATTGATTGSC 361	
QY	764 CTATTGAGAGCAAGAGTATTAGATATGCTATTTTGGCAAGAGAGCTTAAGGAAA 823	
DB	362 CTATTGAGAGCAAGAGTATTAGATATGCTATTTTGGCAAGAGAGCTTAAGGAAA 421	
QY	824 TAAACTTTTGGTTGCAATTTGATGATGCTCTCACCATGGCCACATTTATGATCAG 883	
DB	422 TAAACTTTTGGTTGCAATTTGATGATGCTCTCACCATGGCCACATTTATGATCAG 481	
QY	884 GAGACCAAAAGAAATATATCTTATGATGAAAGATGCTATTGGCATAGTTATTAA 943	
DB	482 GAGACCAAAAGAAATATATCTTATGATGAAAGATGCTATTGGCATAGTTATTAA 541	
QY	944 AGAAAGTGTATTGAGTGGGCTAATCTCAGAAAGGGCTGTTCAGAGCAGCGTGT 1003	
DB	542 AGAAAGTGTATTGAGTGGGCTAATCTCAGAAAGGGCTGTTCAGAGCAGCGTGT 601	
QY	1004 CTTCCTTTAAACTGGATGGAATGAGTACGTGATGATCAGCAAGCTAGCAGCTTAG 1063	
DB	602 CTTCCTTTAAACTGGATGGAATGAGTACGTGATGATCAGCAAGCTAGCAGCTTAG 661	
QY	1064 ATGATCGGAGAAAGAAATGGGCTGTGCTGGAAGAGAGTGGCATATCTTGGAAATGAAG 1123	
DB	662 ATGATCGGAGAAAGAAATGGGCTGTGCTGGAAGAGAGTGGCATATCTTGGTAATTT 721	
QY	1124 TGTCTGATGAAGAGTG 1139	
DB	722 TACTGTAGGAATG 737	
RESULT 13		
ID	ADB83156 standard; cDNA; 789 BP.	
XX	AC ADB83156;	
XX	04-DEC-2003 (first entry)	
XX	Human cDNA sequence useful for the treatment of cancer (SegID 1363).	
XX	human; prostate; cancer; cytostatic; gene therapy; vaccine;	
XX	immune response; gene; ss.	
OS	Homo sapiens.	
XX	WO2003050236-A2.	
XX	19-JUN-2003.	
XX	04-SEP-2002; 2002WO-US028214.	
XX	07-DEC-2001; 2001US-00012697.	
XX	(CHIR) CHIRON CORP.	
XX	(HYSE-) HYSEQ INC.	
XX	Escobedo J, Garcia PD, Kassam A, Jamson G, Drmanac R;	
XX	Crkvenjakov R, Dickson M, Drmanac S, Labat I, Leshkowitz D, Kita D;	
XX	Garcia V, Jones LW, Stache-Crain B, Scott EM;	
XX	WPI; 2003-513972/48.	
XX	New polynucleotides derived from human prostate, useful for modulating immune response to prevent or treat cancer.	
XX	Claim 1; SEQ ID NO 1369; 188pp; English.	
XX	This invention relates to novel isolated polynucleotides of human origin, particularly isolated from the human prostate. Specifically, it refers to the diagnostics and therapeutics comprising these novel human polynucleotides, and includes the derived probes, antisense oligonucleotides and antibodies thereof. The identification of these human prostate genes that can inhibit tumour growth is useful for understanding the progression and nature of complex diseases such as cancer, and hence they are important in the drug discovery process. The present invention describes these polynucleotides and encoded polypeptides as exhibiting cytostatic activity, and through gene therapy and/or vaccines they can be used to modulate the immune response for the prevention or treatment of cancers, particularly of the prostate, but also for breast, lung and colon cancer. This polynucleotide sequence is a human cDNA sequence useful for the treatment of cancer, used in an exemplification of the invention. NOTE: These sequences are not given in the specification but are provided on the WIPO website.	
XX	Sequence 789 BP; 235 A; 131 C; 199 G; 216 T; 0 U; 8 Other;	
SQ	Query Match 49.6%; Score 647.6; DB 8; Length 789; Best Local Similarity 94.8%; Pred. No. 4.5e-173; Matches 729; Conservative 5; Mismatches 22; Indels 13; Gaps 6;	
QY	15C GGGCCGGGAGGAGCAGCAAGCATCCCTCAGAGACATTAAAGCCTGGGGGGTCCC 209	
DB	14 GGGCCGGGAGGAGCAGCAAGCATCCCTCAGAGACATTAAAGCCTGGGGGGTCCC 73	
QY	21C GCTCATTTGGTGGTCTCGGTGCGGCTCGGATTCAGGGGCTTCCAGAGTGTATGGGT 269	
DB	74 GCTCATTTGGTGGTCTCGGTGCGGCTCGGATTCAGGGGCTTCCAGAGTGTATGGGT 133	

QY 270 TTGACAGACCATGATGAAATGAGAACTGTGCGCAAAACAATTTGGTGGCAAGTTCATCG 329
 DB 134 TTGACAGACCATGATGAAATGAGAACTGTGCGCAAAACAATTTGGTGGCAAGTTCATCG 193
 QY 330 AAGAGTTCTGAGTTTCAAAAGACAGCTCTACCTCAGTACAGTGCATCATAGATTTCT 389
 DB 194 AAGAGTTCTGAGTTTCAAAAGACAGCTCTACCTCAGTACAGTGCATCATAGATTTCT 253
 QY 390 TAAATATATATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 449
 DB 254 TAAATATATATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 313
 QY 450 TCCACTGATCTTCAAAAGTTGCAAGATGTTGCAAGAAAGATGATGATGATGATGATGAT 509
 DB 314 TCCACTGATCTTCAAAAGTTGCAAGATGTTGCAAGAAAGATGATGATGATGATGATGAT 373
 QY 510 CTCTGTTGTGAGCGCCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTGCTGAGT 569
 DB 374 CTCGTTGTGAGCGCCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTGCTGAGT 433
 QY 570 GACCGAACCCTCTGAATTTAAATCCAGCTTAAACGGCTCTGTCGACAAAGCTGGATGGAGA 629
 DB 434 GACCGAACCCTCTGAATTTAAATCCAGCTTAAACGGCTCTGTCGACAAAGCTGGATGGAGA 493
 QY 630 ATTATATGAAATGCTCTATTTTATTTGCTTAAAGACATTTGATAGATGCTGATTT 689
 DB 494 ATTATATGAAATGCTCTATTTTATTTGCTTAAAGACATTTGATAGATGCTGATTT 553
 QY 690 GCAGGTGG-AAAATGGCATACTACGAAATGGAGTGGAGTGGAGTGGATAGATG 748
 DB 554 GCAGGTGG-AAAATGGCATACTACGAAATGGAGTGGAGTGGAGTGGATAGATG 612
 QY 749 TGGATATTGATTTGGCTATTGTCAGAGCAAGAGTATTAGATATGCTATTTTGGCAAG 808
 DB 613 TGGATATTGATTTGGCTATTGTCAGAGCAAGAGTATTAGATATGCTATTTTGGCAAG 672
 QY 809 AGAGCTTAAGGAAATTAACCTTTTGGTTGCAATATTGATGATGATGATGATGATGATG 868
 DB 673 AGAGCTTAAGGAA- -TAACTTTTGGTT- -GCATATTGATGATGATGATGATGATGAT 728
 QY 869 ACATTATGATATCAGGACCAAAAGAAATAATATCTATGATGATAA 917
 DB 729 A- - -TTTGTGTCAGGACCA- - -AGATATTGTCATGATGATAA 770

RESULT 14
 ABQ89138
 ID ABQ89138 standard; cDNA; 838 BP.
 AC ABQ89138;
 DT 27-SEP-2002 (first entry)
 DE Human prostate expressed polynucleotide SEQ ID NO 394.
 KW Human; prostate; cytostatic; tumour; cancer; vaccine; gene therapy; gene;
 KW ss.
 OS Homo sapiens.
 EN W0200255700-A2.
 PD 18-JUL-2002.
 PF 07-DEC-2001; 2001WC-US047349.
 XX 07-DEC-2000; 2000US-0254648P.
 XX 13-MAR-2001; 2001US-0275688P.
 XX (CHIR) CHIRON CORP.
 XX (HYSE-) HYSEQ INC.
 PI Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;

PI Crkvenjakov R, Dickson M, Drmanac S, Labat I, Ieshkowitz D, Kita D;
 PI Garcia V, Jones WL, Stache-Crain B, Scott EM;
 XX MPI; 2002-557824/59.
 XX New genes and gene products isolated from human prostate, useful for
 PT treating or diagnosing tumor or cancer (e.g. prostate cancer or breast
 PT cancer), or as vaccines for treating or preventing these diseases.
 XX Claim 1; SEQ ID NO 394; 186pp + Sequence Listing; English.
 XX The invention relates to an isolated polynucleotide comprising any of
 CC 1477 sequences or its fragment, degenerate variant, antisense or
 CC complement. The polynucleotides and gene products are useful for treating
 CC or diagnosing tumor or cancer (e.g. prostate cancer, breast cancer, lung
 CC cancer or medullary carcinoma) in a subject (e.g. cattle, dogs, cats,
 CC rabbits, horse or human). The polynucleotides and polypeptides are also
 CC useful as vaccines for treating or preventing these diseases. The
 CC polynucleotides are useful for gene therapy. The present sequence is that
 CC of one of a group of polynucleotides (ABQ8745-ABQ9015) disclosed
 CC electronically as sequences of the invention. However only 1271
 CC polynucleotide sequences are given, whereas 1477 polynucleotides and 91
 CC proteins are claimed. Note: The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pot_sequence
 XX
 SQ Sequence 838 BP; 241 A; 151 C; 199 G; 227 T; 0 U; 20 Other;
 Query Match 48.8%; Score 636.6; DB 6; Length 838;
 Best Local Similarity 94.7%; Pred. No. 6.2e-170;
 Matches 715; Conservative 4; Mismatches 26; Indels 10; Gaps 6;
 QY 150 GGCCTGGGAGGACAGCAAGGATCCCTCTGAAGAACATTAAAGCAGCTGGCGGGGTCCC 209
 DB 71 GGCCTGGGAGGACAGCAAGGATCCCTCTGAAGAACATTAAAGCAGCTGGCGGGGTCCC 130
 QY 210 GCTCATTTGGTGGTTCCTGGTGGGCTGGATTCAGGGGCTTCCAGAGTGTATGGGT 269
 DB 131 GCTCATTTGGTGGTTCCTGGTGGGCTGGATTCAGGGGCTTCCAGAGTGTATGGGT 190
 QY 270 TTCGACAGACCATGATGAAATGAGAACTGTGCGCAAAACAATTTGGTGGCAAGTTCATCG 329
 DB 191 TTCGACAGACCATGATGAAATGAGAACTGTGCGCAAAACAATTTGGTGGCAAGTTCATCG 250
 QY 330 AAGAGTTCTGAGTTTCAAAAGACAGCTCTACCTCAGTACAGTGCATCATAGATTTCT 389
 DB 251 AAGAGTTCTGAGTTTCAAAAGACAGCTCTACCTCAGTACAGTGCATCATAGATTTCT 310
 QY 390 TAAATATATATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 449
 DB 311 TAAATATATATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 370
 QY 450 TCCTACTGATCTTCAAAAGTTGCAAGATGTTGCAAGAAAGATGATGATGATGATGATGAT 509
 DB 371 TCCTACTGATCTTCAAAAGTTGCAAGATGTTGCAAGAAAGATGATGATGATGATGATGAT 430
 QY 510 CTCTGTTGTGAGCGCCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTGCTGAGT 569
 DB 431 CTCTGTTGTGAGCGCCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTGCTGAGT 490
 QY 570 GACCGAACCCTCTGAATTTAAATCCAGCTTAAACGGCTCTGTCGACAAAGCTGGATGGAGA 629
 DB 491 GACCGAACCCTCTGAATTTAAATCCAGCTTAAACGGCTCTGTCGACAAAGCTGGATGGAGA 550
 QY 630 ATTATATGAAATGCTCTATTTTATTTGCTTAAAGACATTTGATAGATGCTGATTT 689
 DB 551 ATTATATGAAATGCTCTATTTTATTTGCTTAAAGACATTTGATAGATGCTGATTT 610
 QY 690 GCAGGTGG-AAAATGGCATACTACGAAATGGAGTGGAGTGGATAGATGATGATG 748
 DB 611 GCAGGTGG-AAAATGGCATACTACGAAATGGAGTGGAGTGGATAGATGATGATG 669
 QY 749 TGGATATTGATTTGGCTATTGTCAGAGCAAGAGTATTAGATATGCTATTTTGGCAAG 808

Db	670	TGATATGATGCTATTCGAGCAAGAGATTAANATATGGCTTTTGG-TAAG	729
Qy	809	AGAAGCTTAAGCAATAAATCTTTGGTTTCAATATTTGATGATGTCCTCACCATGGCC	868
Db	729	ANAACTTAAG--AAATAACTTTGGTT--GCATATGATGATGTCACCA--TGCG	781
Qy	869	ACATTTATGATCAGGAGACCAAAAAGAAATAATA	903
Db	782	CCNTTTNTGNTCAGGAGACCAAAAAGAAATAATA	816
RESULT 15			
AD382082			
ID	ADB93082	standard; cDNA; 745 BP.	
XX	ADB93082;		
AC	AC		
XX	XX		
DT	04-DEC-2003	(first entry)	
XX	XX		
DE	Human cDNA sequence useful for the treatment of cancer (SeqID 394).		
XX	XX		
KW	humat; prostate; cancer; cytostatic; gene therapy; vaccine;		
KW	immune response; gene; ss.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	WC2003050236-A2.		
XX	XX		
XX	19-JUN-2003.		
XX	XX		
PF	04-SEP-2002; 2002WO-US28214.		
XX	XX		
PR	07-DEC-2001; 2001US-00012697.		
XX	XX		
PA	{CHIR } CHIRON CORP.		
PA	{HYSE-} HYSEQ INC.		
XX	XX		
PI	Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;		
PI	Crkvenjakov R, Dickson M, Drmanac S, Labat I, Leshkowitz D, Kita D;		
PI	Garcia V, Jones LW, Stache-Crain B, Scott EM;		
XX	XX		
DR	WPI; 2003-513972/48.		
XX	XX		
PT	New polynucleotides derived from human prostate, useful for modulating		
PT	immune response to prevent or treat cancer.		
XX	XX		
PS	Claim 1; SEQ ID NO 394; 188pp; English.		
XX	XX		
CC	This invention relates to novel isolated polynucleotides of human origin,		
CC	particularly isolated from the human prostate. Specifically, it refers to		
CC	the diagnostics and therapeutics comprising these novel human		
CC	polynucleotides, and includes the derived probes, antisense		
CC	oligonucleotides and antibodies thereof. The identification of these		
CC	human prostate genes that can inhibit tumour growth is useful for		
CC	understanding the progression and nature of complex diseases such as		
CC	cancer, and hence they are important in the drug discovery process. The		
CC	present invention describes these polynucleotides and encoded		
CC	polypeptides as exhibiting cytostatic activity, and through gene therapy		
CC	and/or vaccines they can be used to modulate the immune response for the		
CC	prevention or treatment of cancers, particularly of the prostate, but		
CC	also for breast, lung and colon cancer. This polynucleotide sequence is a		
CC	human cDNA sequence useful for the treatment of cancer, used in an		
CC	exemplification of the invention. NOTE: These sequences are not given in		
CC	the specification but are provided on the WIPO website.		
XX	XX		
SQ	Sequence 745 BP; 216 A; 129 C; 185 G; 204 T; 0 U; 11 Other;		
Query Match	48.1%;	Score 627.4; DB 8; Length 745;	
Best Local Similarity	95.0%;	Pred. No. 2.4e-167;	
Matches	704; Conservative	4; Mismatches 17; Indels 10; Gaps 6	
QY	150	GGCCCGGGAGGAGCAAGCATCCCGCTCAAGAACATTAAGCACCTGGGGGGGGTCCC	209

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:33 ; Search time 3773.33 Seconds
(without alignments)
10327.779 Million cell updates/sec

Title: US-09-930-440b-3

Perfect score: 1305

Sequence: 1 atgactcgggtggaagg.....ataattcatgcacaaataag 1305

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estha.*

2: em_esthum.*

3: em_estini.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hic.*

9: gb_est1.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1011.2	77.5	1201	9	AL530667
2	927.8	71.1	1201	9	AL542712
3	878	67.3	1088	9	AL533530
4	869.8	66.7	1201	9	AL537974

5	853	65.4	1201	9	AL557407
6	849.8	65.1	1038	9	AL540564
7	830.4	63.6	1201	13	EX406343
8	821.2	62.9	990	9	AL558536
9	801.4	61.4	904	12	BM457667
10	798.6	61.2	881	13	BQ226032
11	786.8	60.3	1201	9	AL538941
12	768.4	58.9	1045	29	AY412329
13	751.6	57.6	982	12	BG674331
14	745.2	57.1	1201	13	EX43798
15	737.8	56.5	874	13	BQ438805
16	727	55.7	1201	9	AL530666
17	722.4	55.4	834	9	AL559436
18	718.4	55.0	805	9	AUL26230
19	718.4	55.0	883	9	AL560379
20	712.4	54.6	823	12	BI837810
21	709.8	54.4	804	12	BI92760
22	708.4	54.3	872	12	BI769833
23	707	54.2	784	12	BI95310
24	706	54.1	884	13	BQ215134
25	705.2	54.0	1036	9	AL545855
26	705	54.0	1087	13	BQ67421
27	703.4	53.9	753	12	BG17075
28	701.4	53.7	802	12	BI97315
29	700.4	53.7	764	14	CA772419
30	699.8	53.6	807	12	BI97820
31	691.6	53.0	912	12	BI462549
32	657	50.3	1201	9	AL569567
33	656	50.3	1201	13	BA463155
34	650	49.8	905	12	BG682092
35	642.6	49.2	732	12	BI600714
36	641	48.1	1201	9	AL537973
37	633.8	48.6	664	12	BM792481
38	633.8	48.6	859	14	CD358582
39	630.8	48.3	1201	13	EX406342
40	624.4	47.8	1045	29	AY412331
41	622	47.7	732	10	BZ253152
42	610.6	46.8	651	10	BZ253161
43	609.8	46.7	636	14	CD672433
44	609.4	46.7	879	14	CD511863
45	609	46.7	1044	9	AL565152

ALIGNMENTS

RESULT 1
AL530667
LOCUS 1201 bp mRNA linear EST 23-MAY-2003
DEFINITION AL530667 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
ACCESSION AL530667
VERSION AL530667.2 GI:31068500
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 13, 2001 this sequence version replaced gi:12794160.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91036 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6598.f For more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODD008BD06021&cluster=6598.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :


```

115 ATGGACTCGTGGAGAGGGGGCCGCCACCTCCGCTCTCCACCCGGGGGGCGACCGTCC 174
116 |||||
117 |||||
118 |||||
119 |||||
120 CGGCGCGCGCGCGCGAGCTCAGCGCAACTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 120
121 |||||
122 |||||
123 |||||
124 CGGCGCGCGCGCGCGAGCTCAGCGCAACTCTCTCGCGCGCGCGCGCGCGCGCGCGCG 234
125 |||||
126 |||||
127 |||||
128 AAGCCCGCGCACTCGCGAGCCCTAATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
129 |||||
130 |||||
131 AAGCCCGCGCACTCGCGAGCCCTAATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 294
132 |||||
133 |||||
134 AAGACATTAGACCTCGCGGGGGTCCGCTCATCTGGCTGGGTCTGCTGGTCCGCGCG 240
135 |||||
136 |||||
137 AAGACATTAGACCTCGCGGGGGTCCGCTCATCTGGCTGGGTCTGCTGGTCCGCGCG 354
138 |||||
139 |||||
140 GATTGAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCGAGAAATG 300
141 |||||
142 |||||
143 GATTGAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCGAGAAATG 414
144 |||||
145 |||||
146 GCCAAACAATTTGTGCAAGTTCATCGAAGAAGTTCTGAAGTTTCAAAGACAGCTCT 360
147 |||||
148 |||||
149 GCCAAACAATTTGTGCAAGTTCATCGAAGAAGTTCTGAAGTTTCAAAGACAGCTCT 474
150 |||||
151 |||||
152 ACCTCAGTAGATGCATCATAGAAATTTCTTAATATATATATGAGGTGACATCTAGGA 420
153 |||||
154 |||||
155 ACCTCAGTAGATGCATCATAGAAATTTCTTAATATATATATGAGGTGACATCTAGGA 534
156 |||||
157 |||||
158 AATATTCAAAGTCTCTTCCTCAAGTTTACATCTACTGATCTTCAAAAGATTGCGAAATG 480
159 |||||
160 |||||
161 AATATTCAAAGTCTCTTCCTCAAGTTTACATCTACTGATCTTCAAAAGATTGCGAAATG 594
162 |||||
163 |||||
164 ATTCGAGAGAGATATGATCTGTTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTG 540
165 |||||
166 |||||
167 ATTCGAGAGAGATATGATCTGTTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTG 654
168 |||||
169 |||||
170 AGTCAAAATTCGAGAGAGATGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 600
171 |||||
172 |||||
173 AGTCAAAATTCGAGAGAGATGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 714
174 |||||
175 |||||
176 CGGCTCGTGCACAAAGCTGGATGGAGAAATATATGAAATGGCTCATTTTATTTTCT 660
177 |||||
178 |||||
179 CGGCTCGTGCACAAAGCTGGATGGAGAAATATATGAAATGGCTCATTTTATTTTCT 774
180 |||||
181 |||||
182 AAAAGACATTTGATAGATGGGTACTTGCAGGGTGG-AAAATGGCATACAGAAATG 719
183 |||||
184 |||||
185 AAAAGACATTTGATAGATGGGTACTTGCAGGGTGGAAAATGGCATACATGAAATG 834
186 |||||
187 |||||
188 CGAGCTGGAACATAGTGTGGATATAGATGTGGATATGATTTGGCTATTTCAGAGCAAG 779
189 |||||
190 |||||
191 CGAGCT-GAACATAGTGTGGATATAGATGTGGATATGATTTGGCTATTTCAGAGCAAG 893
192 |||||
193 |||||
194 AGTATTAGATATGGCTATTGTTGCAAGAGAGCTTAAGGAAATATAAACTTTTGGTTG 839
195 |||||
196 |||||
197 AGTATTAGATATGGCTATTGTTGCAAGAGAGCTTAAGGAAATATAAACTTTTGGTTG 952
198 |||||
199 |||||
200 CAATATTGATGGATCTCCCAATGGCCACA-TTATGATATCAGAGACCAAAAGAA 898
201 |||||
202 |||||
203 CAATATTGATGGATCTCCCAATGGCCACA-TTATGATATCAGAGACCAAAAGAA 1012
204 |||||
205 |||||
206 TAATATCTTATGATTAAGATGCTATTGGGATAAGTTTATTAAGAAAGATGCTATTG 958
207 |||||
208 |||||
209 TAATATCTTATGATTAAGATGCTATTGGGATAAGTTTATTAAGAAAGATGCTATTG 1072
210 |||||
211 |||||
212 AGGTGAGGCTAATCTCAAGAAAGGCGCTGTTTCAAGAGCAGACGCTGCTCTTTTAAACTGG 1018
213 |||||
214 |||||
215 AGGTGAGGCTAATCTCAAGAAAGG--CTGTTTCAAGAGCAGACGCTGCTCTTTTAAACTGG 1129
216 |||||
217 |||||
218 ATTGCAAAATGGAGTCACTGATATCAGACAGAGTATGATGATGATGATGATGATGATG 1075
219 |||||
220 |||||
221 ATTGCAAAATGGAGTCACTGATATCAGACAGAGTATGATGATGATGATGATGATGATG 1186

```

RESULT 3

```

AL533530 1388 bp mRNA linear EST 31-MAY-2003
LOCUS AL533530 Homo sapiens ADULT BRAIN Homo sapiens cDNA clone
DEFINITION CS0DN004YA17 5-PRIME, mRNA sequence.
ACCESSION AL533530
VERSION AL533530.2 GI:31260610
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1088)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 13, 2001 this sequence version replaced gi:12737023.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6598.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DN004AA09QP1&cluster=6598.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGer Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DN004AA09QP1.
FEATURES
source
1..1088
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DN004YA17"
/tissue_type="ADULT BRAIN"
/dev_stage="adult"
/clone_lib="Homo sapiens ADULT BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
ORIGIN
Query Match 67.3%; Score 878; DB 9; Length 1088;
Best Local Similarity 97.6%; Pred. No. 8.3e-220;
Matches 943; Conservative 10; Mismatches 5; Indels 8; Gaps 6;
QY 1 ATGGACTCGTGGAGAGGGGGCCGCCACCTCCGCTCTCCACCCGGGGGGCGACCGTCC 60
DB 79 ATGGACTCGTGGAGAGAGGGGGCCGCCACCTCCGCTCTCCACCCGGGGGGCGACCGTCC 138
QY 61 CGGCGCGCGCGCGCGAGCTCAGCGCAACTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 120
DB 139 CGGCGCGCGCGCGCGAGCTCAGCGCAACTCTCGCGCGCGCGCGCGCGCGCGCGCG 198
QY 121 AAGCCCGCGCACTCGCGAGCCCTAATCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
DB 199 AAGCCCGCGCACTCGCGAGCCCTAATCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 258
QY 181 AAGACATTAGACCTCGCGGGGGTCCGCTCATCTGGTCTGGTCTGGTCTGGTCTGG 240
DB 259 AAGACATTAGACCTCGCGGGGGTCCGCTCATCTGGTCTGGTCTGGTCTGGTCTGG 318
QY 241 GATTGAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCGAGAAATG 300
DB 319 GATTGAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCGAGAAATG 378
QY 301 GCCAAACAATTTGTGCAAGTTCATCGAAGAAGTTCTGAAGTTTCAAAGACAGCTCT 360
DB 379 GCCAAACAATTTGTGCAAGTTCATCGAAGAAGTTCTGAAGTTTCAAAGACAGCTCT 438
QY 361 ACCTCAGTAGATGCATCATAGAAATTTCTTAATATATATATGAGGTGACATCTAGGA 420

```



```

COMMENT      On Feb 15, 2001 this sequence version replaced gi:12870832.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6598.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DE002DA01QP1&cluster=6598.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DE002DA01QP1.

FEATURES
source
1. .1038
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE002YB02"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match      65.1%; Score 849.8; DB 9; Length 1038;
Best Local Similarity 96.1%; Pred. No. 2.2e-212;
Matches 889; Conservative 20; Mismatches 12; Indels 4; Gaps 4;

QY 1 ATGGACTCGTGGAGAGAGGGGGCGCCACCTCGCTCCACCCGGGGGGCGACCGTCC 60
DB 115 ATGGACTCGTGGAGAGAGGGGGCGCCACCTCGCTCCACCCGGGGGGCGACCGTCC 174
QY 61 CGGGGGCGCGCGCGAGCTGCAGCGCACTCTCGCGGGGGCGAGCGCGAGTGGAG 120
DB 175 CGGGGGCGCGCGCGAGCTGCAGCGCACTCTCGCGGGGGCGAGCGCGAGTGGAG 234
QY 121 AAGCCCCCGCACCTGCGAGCCCTTAATCTTGGCCCCGGGGAGCGCAAGGCACTCC 180
DB 235 AAGCCCCCGCACCTGCGAGCCCTTAATCTTGGCCCCGGGGAGCGCAAGGCACTCC 294
QY 181 AAGCAATTAAGCACTCGCGGGGGTCCCGCTCATTTGGCTGGCTCGTGGGGCCCTG 240
DB 295 AAGCAATTAAGCACTCGCGGGGGTCCCGCTCATTTGGCTGGCTCGTGGGGCCCTG 354
QY 241 GATTCAGGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCAGAAATG 300
DB 355 GATTCAGGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCAGAAATG 414
QY 301 GCGCAACAATTTGGTGCACAAGTTTCAGAGAGTTCGAAGTTTCAAAAGCAGCTCT 360
DB 415 GCGCAACAATTTGGTGCACAAGTTTCAGAGAGTTCGAAGTTTCAAAAGCAGCTCT 474
QY 361 ACCTCACTAGATGCCATCATAGAAATTCCTAATTAATTAATCAGCTGACATGTAGGA 420
DB 475 ACCTCACTAGATGCCATCATAGAAATTCCTAATTAATTAATCAGCTGACATGTAGGA 534
QY 421 AATATTCAAGTACTTCTTCATGTTTACATCTTACTGATCTTCAAAAAGTTCAGAAATG 480
DB 535 AATATTCAAGTACTTCTTCATGTTTACATCTTACTGATCTTCAAAAAGTTCAGAAATG 594
QY 481 ATTGAGAGAGAGGATATGATTCGTTCTCTGTTGAGAGCGCATCATGTTTCGATGG 540
DB 595 ATTGAGAGAGAGGATATGATTCGTTCTCTGTTGAGAGCGCATCATGTTTCGATGG 654
QY 541 AGTGAATTCAGAAAGAGTTCGTTGAGTGCACCGACCTCTGATTTAAATCCAGCTAA 600
DB 655 AGTGAATTCAGAAAGAGTTCGTTGAGTGCACCGACCTCTGATTTAAATCCAGCTAA 714
QY 601 CGGCTCTCGTGCAGAGACTGGGATGGAGAAATTAATGAAATGGCTCAATTTATTTGCT 660

```

```

DB 715 CGGCCCTCGTGCAGACACTGGGATGGAGATTAATGAAATGGCTCATTTATTTGCT 774
QY 661 AAAAGACATTTGATAGAGATGGGTTACTTGCAGGGTGG-AAAATGGCATCTACGAATG 719
DB 775 AAAAGACATTTGATAGAGATGGGTTACTTGCAGGGTGG-AAAATGGCATCTACGAATG 834
QY 720 CGAGCTGGAGACATAGTGTGATATAGATATGATATGATTTGGCTTATTCAGAGCAAG 779
DB 835 CGAGCT-GAACAATAGTGTGATATAGATATGATTTGGCTTATTCAGAGCAAG 893
QY 780 AGTATTAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGCAATATAAAC-TTTTGGTTT 838
DB 894 AGTATTAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGCAATATAAAC-TTTTGGTTT 953
QY 839 GCAATATTGATGATGTCTCACCAGATGGCCACAT-TTATGATCAGAGAGCAAAAGAA 897
DB 954 GCRATATTGATGATGTCTCTCCATAGGCMCATATTTTGTTCAGGGGGCCAAAAAAA 1013
QY 898 ATAATATCTTATGATGTAAAGATG 922
DB 1014 AAAAATCTTTTGTGTGTAAAAAG 1038

RESULT 7
LOCUS      BX406343      1201 bp      mRNA      linear      EST 15-MAY-2003
DEFINITION BX406343 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
ACCESSION . BX406343
VERSION    BX406343.1 GI:30764679
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE   1 (bases 1 to 1201)
AUTHORS    Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT     Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6598.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0AM011BD06QP1&cluster=6598.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0AM011BD06QP1.

FEATURES
source
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM011YG12"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match      63.6%; Score 830.4; DB 13; Length 1201;
Best Local Similarity 93.3%; Pred. No. 3e-207;
Matches 940; Conservative 21; Mismatches 33; Indels 13; Gaps 9;

QY 1 ATGGACTCGTGGAGAGAGGGGGCGCCACCTCGCTCTCCAAACCCGGGGGGCGACCGTCC 60

```

```

160 ATGACCTCGGTGAGAAAGGGGGCGGCACCTCGTCTCCAAACCGCGGGGCGACCGTCC 219
QY
61 CGGGGGCGGGCGGCGAGCTGACGCAAACTCTCGCGCGGGCCAGGCGCGAGGCTGTGGAG 120
Db
220 CGGGGGCGGGCGGCGAGCTGACGCAAACTCTCGCGCGGGCCAGGCGCGAGGCTGTGGAG 279
QY
121 AAGCCCCCGCACCTGGCAGCCCTAAATCTGTGGCGGGGAGGAGCAAAAGGATCCCCCTG 180
Db
280 AAGCCCCCGCACCTGGCAGCCCTAAATCTGTGGCGGGGAGGAGCAAAAGGATCCCCCTG 339
QY
181 AAGAACATTAAGACACTGGCGGGGCTCCGCTCATTTGGCTGGCTCGCTGGCGGCGCTG 240
Db
340 AAGAACATTAAGACACTGGCGGGGCTCCGCTCATTTGGCTGGCTCGCTGGCGGCGCTG 399
QY
241 GATTCAGGGGCGCTTCAGAGCTGTATGGTTCAGACAGACCATGATGAATTTGAGAATGTG 300
Db
400 GATTCAGGGGCGCTTCAGAGCTGTATGGTTCAGACAGACCATGATGAATTTGAGAATGTG 459
QY
301 GCCAAACAATTTGGTGCAAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTCT 360
Db
460 GCCAAACAATTTGGTGCAAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTCT 519
QY
361 ACCTCACTAGATGCATCATAGAAATTTCTAAATTAATATATAGAGTGTGACATTTGTAGGA 420
Db
520 ACCCTCACTAGATGCATCATAGAAATTTCTAAATTAATATATAGAGTGTGACATTTGTAGGA 579
QY
421 AATATTCAAGCTACTCTTCATGCTTTACATCTCTACTGATCTTCAAAAGTTTCAGAAATG 480
Db
580 AATATTCAAGCTACTCTTCATGCTTTACATCTCTACTGATCTTCAAAAGTTTCAGAAATG 639
QY
481 ATTCGAGAAGAGATATGATTTCTGTTTCTGTTGTGAGAGCCCATGATTTGATGATG 540
Db
640 ATTCGAGAAGAGATATGATTTCTGTTTCTGTTGTGAGAGCCCATGATTTGATGATG 699
QY
541 AGTGAATTCAGAAAGAGTTCGTGAGTACGAACTCTGATTTAAATTCAGAGCTAAA 600
Db
700 AGTGAATTCAGAAAGAGTTCGTGAGTACGAACTCTGATTTAAATTCAGAGCTAAA 759
QY
601 CGGCGCTCGTCGCAAGACTGGGATGAGAAATTAATGAAATGGCTCTATTTATTTGCT 660
Db
760 CGGCGCTCGTCGCAAGACTGGGATGAGAAATTAATGAAATGGCTCTATTTATTTGCT 819
QY
661 AAGAGCAATTTGATAGAGTGGTTACTTCGAGGCTG-BAATGCGATACACTAGAAATG 719
Db
820 AAGAGCAATTTGATAGAGTGGTTACTTCGAGGCTGAAATTTGCGATACACTAGAAATG 879
QY
720 CGAGCTGGAACATAGTGTGATATAGATGCGATATGATTTGATTTGCTATTCAGAGCAAG 779
Db
880 CGAGC-GGAACATAGTGTGATATAGATGCGATATGATTTGATTTGCTATTCAGAGC-AAG 936
QY
780 AGTATTAAGATATGGCTATTTTGGCAAGAGCTTTAAGAAATAAAATCTTTGGTTG 839
Db
937 AGTATTAAGATATGGCTATTTTGGCAAGAGC--GCTTAGAAATAACACTTTTGGGTTK 994
QY
840 CAATTTGATGATGCTTCCACATGCGCATTTATGATTCAGAGCAAAAGAAAT 899
Db
995 CNAATTTGATGATGCTTCCACATGCGCATTTATGATTCAGAGCAAAAGAAAT--GAT 1051
QY
900 AATATCTTATGATGATAAGATGCTATTTGGGATAAGTTTATTAAGAAAGAGTGTGATTTGA 959
Db
1052 WATATCTTATGATGATAAGATGCTATTTGGGATA--GTTTATTAAGAAAGTGTGATTTGA 1109
QY
960 GGTGAGGCTAATCTCAGAAAGGGCTGTTTCAAGCAGACAGCTGCTT 1006
Db
1110 G--WGAGGTAATVMTAAAGGGCTTTTCAAGAGMRGMSYGTGTTTTT 1154

```

RESULT 8
 AL558536 990 bp mRNA linear EST 31-MAY-2003
 LOCUS AL558536 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
 DEFINITION Homo sapiens cDNA clone CS0DJ006YB11 5-PRIME, mRNA sequence.

ACCESSION AL558536 GI:31280334
 VERSION AL558536.2
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1. (bases 1 to 990)
 AUTHORS Li, W.B., Graber, C., Jesse, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12903150.
 Contact: Genoscope Centre National de Sequencage
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6598.f For more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DJ006CA06QPI&cluster=6598.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DJ006CA06QPI.
 FEATURES
 Location/Qualifiers
 1..990
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DJ006YB11"
 /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /cell_line="JURKAT"
 /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and Eco3 V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 62.9%; Score 821.2; DB 9; Length 990;
 Best Local Similarity 98.4%; Pred. No. 7.4e-205;
 Matches 872; Conservative 8; Mismatches 1; Indels 5; Gaps 5;
 QY 1 ATGACCTCGGTGAGAAAGGGGGCGGCACCTCGTCTCCAAACCGCGGGGCGACCGTCC 60
 Db 109 ATGACCTCGGTGAGAAAGGGGGCGGCACCTCGTCTCCAAACCGCGGGGCGACCGTCC 168
 QY 61 CGGGGGCGGGCGGCGAGCTGACGCAAACTCTCGCGGGCCAGGCGCGAGGCTGTGGAG 120
 Db 169 CGGGGGCGGGCGGCGAGCTGACGCAAACTCTCGCGGGCCAGGCGCGAGGCTGTGGAG 228
 QY 121 AAGCCCCCGCACCTGGCAGCCCTAAATCTGTGGCGGGGAGGAGCAAAAGGATCCCCCTG 180
 Db 229 AAGCCCCCGCACCTGGCAGCCCTAAATCTGTGGCGGGGAGGAGCAAAAGGATCCCCCTG 288
 QY 181 AAGAACATTAAGACACTGGCGGGGCTCCGCTCATTTGGCTGGCTCGCTGGCGGCGCTG 240
 Db 289 AAGAACATTAAGACACTGGCGGGGCTCCGCTCATTTGGCTGGCTCGCTGGCGGCGCTG 348
 QY 241 GATTCAGGGGCGCTTCAGAGCTGTATGGTTCAGACAGACCATGATGAATTTGAGAATGTG 300
 Db 349 GATTCAGGGGCGCTTCAGAGCTGTATGGTTCAGACAGACCATGATGAATTTGAGAATGTG 408
 QY 301 GCCAAACAATTTGGTGCAAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTCT 360
 Db 409 GCCAAACAATTTGGTGCAAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTCT 468
 QY 361 ACCTCACTAGATGCATCATAGAAATTTCTAAATTAATATATAGAGTGTGACATTTGTAGGA 420
 Db 469 ACCTCACTAGATGCATCATAGAAATTTCTAAATTAATATATAGAGTGTGACATTTGTAGGA 528
 QY 421 AATATTCAAGCTACTCTTCATGCTTTACATCTCTACTGATCTTCAAAAGTTTCAGAAATG 480


```

VERSION B0226032.1 GI:20407432
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC http://mgl.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: L1AM13345 row: i column: 12
High quality sequence stop: 690.
FEATURES
    source
        1..881
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6066899"
            /tissue_type="embryonal carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_92"
            /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
            Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 2.5 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
ORIGIN
Query Match 61.2%; Score 798.6; DB 13; Length 881;
Best Local Similarity 97.8%; Pred. No. 6.3e-199;
Matches 826; Conservative 4; Mismatches 13; Indels 2; Gaps 2;
QY 1 ATGACTCGGTGGAGAGGGGGCGCCACCTCCGCTCCACCCCGGGGGCGACGCC 60
DB 10 ATGACTCGGTGGAGAGGGGGCGCCACCTCCGCTCCACCCCGGGGGCGACGCC 69
QY 61 CGGGGCGGGCGCCGAGCTGCGAGCAACTCTCGCGGGCGCCAGGGCGGAGGTGGAG 120
DB 70 CGGGGCGGGCGCCGAGCTGCGAGCAACTCTCGCGGGCGCCAGGGCGGAGGTGGAG 129
QY 121 AAGCCCCCGACCTGGCGAGCCCTTAATCTGGCCCGGGGAGGAGCAAGAGCATCCCGCTG 180
DB 130 AAGCCCCCGACCTGGCGAGCCCTTAATCTGGCCCGGGGAGGAGCAAGAGCATCCCGCTG 189
QY 181 AAGAACATTAGACCTCGCGGGGGTCCCGCTCATTTGGCTGGTCTCGGTGGCGCCCTG 240
DB 190 AAGAACATTAGACCTCGCGGGGGTCCCGCTCATTTGGCTGGTCTCGGTGGCGCCCTG 249
QY 241 GATTTCAGGGGCTTCCAGAGGTGTAGGGTTTCGACAGACCATGATGAATTCAGAAATGTG 300
DB 250 GATTTCAGGGGCTTCCAGAGGTGTAGGGTTTCGACAGACCATGATGAATTCAGAAATGTG 309
QY 301 GCCAAACAAATTTGGTGCACAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTCT 360
DB 310 GCCAAACAAATTTGGTGCACAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTCT 369
QY 361 ACCCTACCTAGATGCATCATGAATTTCTTAATTAATTAATGAGGTGCATTTGAGA 420
DB 370 ACCCTACCTAGATGCATCATGAATTTCTTAATTAATTAATGAGGTGCATTTGAGA 429
QY 421 AATATTCAAGCTACTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTCGAGAAATG 480
DB 430 AATATTCAAGCTACTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTCGAGAAATG 489

```

```

QY 481 ATTGAGAAAGAGHATGATTCCTGTTTCTCTGTTTGGAGACGCGATCAGTTTCGATGG 540
DB 490 ATTGAGAAAGAGGATATGATTCCTGTTTCTCTGTTTGGAGACGCGATCAGTTTCGATGG 549
QY 541 AGTGAATTTGAGAAAGAGTTCGTCAGTGCACCGAACCTCTGAATTTAAATCCAGCTAAA 600
DB 550 AGTGAATTTGAGAAAGAGTTCGTCAGTGCACCGAACCTCTGAATTTAAATCCAGCTAAA 609
QY 601 CGGCTCTGTCAGCAAGACTGGGATGGAGAAATATATAGAAATGGCTCATTTTATTTGCT 660
DB 610 CGGCTCTGTCAGCAAGACTGGGATGGAGAAATATATAGAAATGGCTCATTTTATTTGCT 669
QY 661 AAAGACATTTGATAGATGGTTCCTGAGGGTGG-AAAATGGCTACTACGAAATG 719
DB 670 AAAGACATTTGATAGAAATGGTTCCTGAGGGTGGAAATATGATTTGCTACTACGAAATG 729
QY 720 CGAGCTGGAACATAGTGTGGATATAGATGT-GGATATTGATTGGCTATTGCGAGCAAA 778
DB 730 CGAGCTGGAACATAGTGTGGATATAGATGTGGATATTGATTGGCTATTGCGAGCAAA 789
QY 779 GAGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAGGAATATAAATCTTTGGTTT 838
DB 790 GAGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAGGAATATAAATCTTTGGTTT 849
QY 839 GCAAT 843
DB 850 GCAAT 854

```

```

RESULT 11
AL538941
LOCUS
DEFINITION
    AL538941 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
    CS0DF030YB08 5-PRIME, mRNA sequence.
ACCESSION
    AL538941
VERSION
    AL538941.2 GI:31263511
KEYWORDS
    EST.
SOURCE
    Homo sapiens (human)
ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE
    1 (bases 1 to 1201)
    Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
    Full-length cDNA libraries and normalization
    Unpublished (2001)
JOURNAL
    On Feb 15, 2001 this sequence version replaced gi:12867704.
COMMENT
    Contact: Genoscope
    Genoscope - Centre National de Sequencage
    BP 191 91006 EVRY cedex - France
    Email: segret@genoscope.cns.fr, Web: www.genoscope.cns.fr
    Library was constructed by Life Technologies, a division of
    Invitrogen. This sequence belongs to sequence cluster 6598.f For
    more information about this cluster, see
    http://www.genoscope.cns.fr/
    cgi-bin/cluster.cgi?seq=CS0DF030A04QP1&cluster=6598.f. Contact :
    Feng Liang Email : fliang@lifetech.com URL :
    http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
    Faraday Avenue Genoscope sequence ID : CS0DF030DA04QP1.

```

```

FEATURES
    source
        1..1201
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DF030YB08"
            /tissue_type="FETAL BRAIN"
            /dev stage="fetal"
            /clone_lib="Homo sapiens FETAL BRAIN"
            /note="Organ: Brain; Vector: pCMVSPORT_6; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-strand cDNA was digested with Not I and
            cloned into the Not I and EcoRV sites of the pCMVSPORT 6
            vector. Library was not normalized."

```

ORIGIN

QY 801 TGCAGAGAGAGCTTAAGAGAAATAAACTTTTGGTTGCAATATTGATGATGCTCTCAC 860
 Db 541 NNN 600
 QY 861 CAATGGCCACATTTATGATCAGGAGACCAAAAGAACTAATATCTTATGATGTAAAGA 920
 Db 601 NNN 660
 QY 921 TGCTATTGGGATAAGTTTATTAAGAAAGTGTGTTATTAAGTGGCTTAATCTCAGAAAG 980
 Db 661 NNN 720
 QY 981 GGCTCTTAAAGCAGCGCTGCTCTTTTAAACTGAGTTGCAAAATGGAAGTCAAGTGT 1040
 Db 721 GGCTCTTAAAGCAGCGCTGCTCTTTTAAACTGAGTTGCAAAATGGAAGTCAAGTGT 780
 QY 1041 ATCAGACAAAGCTAGCTGATGATGAATCGAGAAAGAAATGGGCTGTGCTGGAAGA 1100
 Db 781 ATCAGACAAAGCTAGCTGATGATGAATCGAGAAAGAAATGGGCTGTGCTGGAAGA 840
 QY 1101 AGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTGTCTGAGAGAGTGGGCTTAAG 1160
 Db 841 AGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTGTCTGAGAGAGTGGGCTTAAG 900
 QY 1161 TGGCGCTCTGCTGATGCTGCTCTTACGCGCCAGAGGCTGTGATGATGATGCAATG 1220
 Db 901 TGGCGCTCTGCTGATGCTGCTCTTACGCGCCAGAGGCTGTGATGATGATGCAATG 960
 QY 1221 TAAATGCTGGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1280
 Db 961 TAAATGCTGGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 QY 1281 AGTTAATAATTCATGCAAAAATAG 1305
 Db 1021 GGTAAATAATTCATGCAAAAATAG 1045

RESULT 13
 BG674331
 LOCUS
 DEFINITION 602620310F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4745787 5',
 mRNA sequence.
 ACCESSION BG674331
 VERSION BG674331.1 GI:13905727
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 982)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: James Cleaver, M.D.
 cDNA library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA
 Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLM10593 row: c column: 04
 High quality sequence stop: 827.
 Location/Qualifiers
 1..982
 /organism="Homo sapiens"
 /molecule="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4745787"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI_CGAP_Skn3"

FEATURES
 source

/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 ORIGIN
 Query Match 57.6%; Score 751.6; DB 12; Length 982;
 Best Local Similarity 97.7%; Pred. No. 1.6e-186;
 Matches 824; Conservative 1; Mismatches 11; Indels 7; Gaps 6;
 QY 466 AAAGTTTCAGAAATGATTTCAGAGAGAGGATATGATCTCTGTTCTCTGTTGAGAGCGC 525
 Db 2 AAAGTTTCAGAAATGATTTCAGAGAGAGGATATGATCTCTGTTCTCTGTTGAGAGCGC 61
 QY 526 CATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTCGTGAAAGTCAAGCACTCTGAAAT 585
 Db 62 CATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTCGTGAAAGTCAAGCACTCTGAAAT 121
 QY 586 TTAATCCAGCTAAACGGCCCTCGTCGACAGACTGGGATGGAGATTAATATGAAAATGCGC 645
 Db 122 TTAATCCAGCTAAACGGCCCTCGTCGACAGACTGGGATGGAGATTAATATGAAAATGCGC 181
 QY 646 TCATTTTATTTTGTCTAAAGACATTTTCATAGAGATGGTTACTTTCAGGGTGG-AAAATG 704
 Db 182 TCATTTTATTTTGTCTAAAGACATTTTCATAGAGATGGTTACTTTCAGGGTGGAAAATG 241
 QY 705 GCATACACGAAATGCGAGCTGGAAACATAGTGTGGATATAGATGTGGATATGATTTGTTGCC 764
 Db 242 GCATACACGAAATGCGAGCT-GAACAATAGTGTGGATATAGATGTGGATATGATTTGTTGCC 300
 QY 765 TATTCAGACCAAGAGTATTAAGATATGCTATTTTGGCAAGAGAGAGCTTAAGGAAT 824
 Db 301 TATTCAGACCAAGAGTATTAAGATATGCTATTTTGGCAAGAGAGAGCTTAAGGAAT 360
 QY 825 AAAACTTTTGGTTTGGCAATATGATGATGCTCTCACCATGGCCACATTTATGATCAGG 884
 Db 361 AAACTTTTGGTTTGGCAATATGATGATGCTCTCACCATGGCCACATTTATGATCAGG 420
 QY 885 AGACCAAAAGAAATATATCTTATGATGTAAAGATGCTATTGGGATAAGTTTATTTAAA 944
 Db 421 AGACCAAAAGAAATATATCTTATGATGTAAAGATGCTATTGGGATAAGTTTATTTAAA 480
 QY 945 GAAAGTGGTATTGAGTGGAGGCTTAATCTCAGAAAGGCGCTGTTCAAGCAGAGCGCTGC 1004
 Db 481 GAAAGTGGTATTGAGTGGAGGCTTAATCTCAGAAAGGCGCTGTTCAAGCAGAGCGCTGC 540
 QY 1005 TTCTTTAAAACCTGGATTGCAAAATGGAAGTCAAGTGTATCAGCAAGCTAGCAGTTGTAGA 1064
 Db 541 TTCTTTAAAACCTGGATTGCAAAATGGAAGTCAAGTGTATCAGCAAGCTAGCAGTTGTAGA 600
 QY 1065 TGAATCGAGAAAGAAATGGGCTGTGCTGGAAAGAGTGGCATATCTTGGAAATGAAGT 1124
 Db 601 TGAATCGAGAAAGAAATGGGCTGTGCTGGAAAGAGTGG-ATATCTTGGAAATGAAGT 659
 QY 1125 GTCTGATGAGAGTGTCTTCAGAGAGTGGGCTTAAGTGGCGCTCTCTGATGATGCTGCTTC 1184
 Db 660 GTCTGATGAGAGTGTCTTCAGAGAGTGGGCTTAAGTGGCGCTCTCTGATGATGCTGCTTC 729
 QY 1185 --CTAGCCCAAGAGGCTGTGATACATTTGCAAAATGTAATGGTGGCGCTGGTGCATC 1242
 Db 720 TACTGGCCCAAGAGGCTGTGATACATTTGCAATTAATGGTGGCGTGGTGGTGGTGGTGGTGG 778
 QY 1243 CGAGAAATTTGCGAGCAGACATTTGCTTAAATGGAAGAAAGTAAATTAATCATGCCAAAAA 1302
 Db 779 CGAGAAATTTGCGAGCAGACATTTG-CTACTAAATGGAAGAAAGTAAATTAATCATGCCAAAA 837
 QY 1303 TAG 1305
 Db 838 TAG 840
 RESULT 14
 BX343798

LOCUS BX343798 1201 bp mRNA linear EST 02-MAY-2003
 DEFINITION BX343798 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0D1023VA10 5-PRIME, mRNA sequence.
 ACCESSION BX343798
 VERSION BX343798.1 GI:30344204
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1201)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6598.f For more information about this cluster, see
 http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS1A1006ZE06QPI&cluster=6598.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Paraday Avenue Genoscope sequence ID : CS1A1006ZE06QPI.
 FEATURES
 source
 1..1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1023VA10"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /notes="1st strand cDNA was primed with a NotI-oligo (dr) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
 ORIGIN
 Query March 57.1%; Score 745.2; DB 13; Length 1201;
 Best Local Similarity 97.7%; Pred. No. 8.5e-185;
 Matches 823; Conservative 6; Mismatches 4; Indels 9; Gaps 7;
 QY 1 ATGACTCGTGGAGAGGGGGCGCCACCTCCCTCTCCACCGCGGGCGGACCTCC 60
 DB 146 ATGACTCGTGGAGAGGGGGCGCCACCTCCCTCTCCACCGCGGGGGCGGACCTCC 205
 QY 61 CGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 120
 DB 206 CGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 265
 QY 121 AAGCCCCCGACCTGGAGGCTTAATCTGGCCCGGGAGGCGAGCAAGGATCCCGCTG 180
 DB 266 AAGCCCCCGACCTGGAGGCTTAATCTGGCCCGGGAGGCGAGCAAGGATCCCGCTG 325
 QY 181 AAGACATTAAGCACTGGCGGGGGTCCGCTCATCTGGCTGGGCTCTGGCTGGCGCCCTG 240
 DB 326 AAGACATTAAGCACTGGCGGGGGTCCGCTCATCTGGCTGGGCTCTGGCTGGCGCCCTG 385
 QY 241 GATTCAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATGAGAATGTG 300
 DB 386 GATTCAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATGAGAATGTG 445
 QY 301 GCCAAACAATTTGGTGCACAGTTTCATCGAAGAGTCTGTGAAGTTTCAAGAAGACAGCTCT 360
 DB 446 GCCAAACAATTTGGTGCACAGTTTCATCGAAGAGTCTGTGAAGTTTCAAGAAGACAGCTCT 505
 QY 361 ACCTCATAGTGCATCAGATTTCTTAATATATATATATATATATATATATATATATATATAT 420
 DB 506 ACCTCATAGTGCATCAGATTTCTTAATATATATATATATATATATATATATATATATATAT 565
 QY 421 AATATTCAGCTACTTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTGCAGAAATG 480

DB 566 AATATTCAGCTACTTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTGCAGAAATG 625
 QY 481 ATTGAGAGAGAGATATGATTCCTGTTTCTCTGTTGTGAGACCCCATCAGTTTCGATGG 540
 DB 626 ATTGAGAGAGAGATATGATTCCTGTTTCTCTGTTGTGAGACCCCATCAGTTTCGATGG 685
 QY 541 APTGAAATTCAGAAAGGAGTTCTGAGTGCACCAACCTCTGAATTTAAATCCAGCTAAA 600
 DB 686 AGTGAATTCAGAAAGGAGTTCTGAGTGCACCAACCTCTGAATTTAAATCCAGCTAAA 745
 QY 601 CGGCTCTGTCGACCAAGACTGGGATGGAATATATATATATATATATATATATATATATAT 660
 DB 746 CGGCTCTGTCGACCAAGACTGGGATGGAATATATATATATATATATATATATATATAT 805
 QY 661 AATGACATTTGATGAGATGGGTACTTTCAGAGGTGG-AAAATGGCATATACGAATG 729
 DB 806 AATGACATTTGATGAGATGGGTACTTTCAGAGGTGG-AAAATGGCATATACGAATG 865
 QY 720 CGAGCTGGAACATAGTGTGATATAGATGTGATATGATGCTTATGCTTATGCTTATGCTT 779
 DB 866 CGAGCT-GAACATAGTGTGATATAGATGTGATATGATGCTTATGCTTATGCTTATGCTT 922
 QY 780 AGTATTAAGTATGCTTATTTTGGCAAGAGAGAGCTTAAAGGAAATAAACTTTTGTGTTG 839
 DB 923 AGTATTAAGTATGCTTATTTTGGCAAGAGAGAGCTTAAAGGAAATAAACTTTTGTGTTG 977
 QY 840 CA 841
 DB 978 CA 979
 RESULT 15
 BQ438805
 LOCUS BQ438805 874 bp mRNA linear EST 24-MAY-2002
 DEFINITION AGENCOURT 7809455 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6053483 5', mRNA sequence.
 ACCESSION BQ438805
 VERSION BQ438805.1 GI:21177881
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 874)
 AUTHORS K-H-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 E-mail: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILML)
 RNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILML at:
 http://image.llnl.gov
 Plate: ILML3310 row: j column: 12
 High quality sequence stop: 602.
 FEATURES
 Location/Qualifiers
 1..874
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6053483"
 /tissue_type="epithelioid carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_70"
 /notes="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."
 ORIGIN

Query Match 56.5%; Score 737.8; DB 13; Length 874;
Best Local Similarity 97.5%; Pred. No. 6.6e-183;
Matches 787; Conservative 4; Mismatches 11; Indels 5; Gaps 4;

Qy	1	ATGACTCGGTGGAGAGGGGCGCCACCTCCCTCTCCAAACCCGCGGGGCGGACGCTCC	50
Db	69	ATGACTCGGTGGAGAGGGGCGCCACCTCCCTCTCCAAACCCGCGGGGCGGACGCTCC	128
Qy	61	CGGGCCCGGCGCCGAGCTCGAGGCAACTCTCGCGCGGCGGCGGCGGAGTGTGGAG	120
Db	129	CGGGCCCGGCGCCGAGCTCGAGGCAACTCTCGCGCGGCGGCGGCGGAGTGTGGAG	188
Qy	121	AAGCCCCCGACCTGGCGAGCCCTAATCTGCGCGGCGGCGGCGGAGGCAAAAGGCAATCC	180
Db	189	AAGCCCCCGACCTGGCGAGCCCTAATCTGCGCGGCGGCGGCGGAGGCAAAAGGCAATCC	248
Qy	181	AAGAACATTAAAGCACTCGCGGGGTCCGCTCATTTGGCTGGGTCTCGTGGCGGCCCTG	240
Db	249	AAGAACATTAAAGCACTCGCGGGGTCCGCTCATTTGGCTGGGTCTCGTGGCGGCCCTG	308
Qy	241	GATTCAGGGGCGCTTCAGAGTGTATGGTTTCGACAGACCATGATGAATTCAGAAATGTG	300
Db	309	GATTCAGGGGCGCTTCAGAGTGTATGGTTTCGACAGACCATGATGAATTCAGAAATGTG	368
Qy	301	GCCAAACAATTTGGTGGCAAGTTCATCGAAGAGTTCGAAGTTTCAAAAGACAGCTCT	360
Db	369	GCCAAACAATTTGGTGGCAAGTTCATCGAAGAGTTCGAAGTTTCAAAAGACAGCTCT	428
Qy	361	ACCTCAGTAGATGACATAGAAATCTTAATATATATATATATATATATATATATATAT	420
Db	429	ACCTCAGTAGATGACATAGAAATCTTAATATATATATATATATATATATATATATAT	488
Qy	421	AATATTCAGCTACT	480
Db	489	AATATTCAGCTACT	548
Qy	481	ATTGAGAGAGGATATGATTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTG	540
Db	549	ATTGAGAGAGGATATGATTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTG	608
Qy	541	AGTGAATTCAGAAAGGAGTTCGTGAGTGACCGAACCTCTGAATTTAAATCCAGCTAA	600
Db	609	AGTGAATTCAGAAAGGAGTTCGTGAGTGACCGAACCTCTGAATTTAAATCCAGCTAA	668
Qy	601	CGGCTCTCGACAGACTGGGATCGAGAAATATATGAAATGGCTCATTTTATTTTGT	660
Db	669	CGGCTCTCGACAGACTGGGATCGAGAAATATATGAAATGGCTCATTTTATTTTGT	728
Qy	661	AAAGACATTTGATAGAGATGGTTTACTTGCAGGTGG-AAAATGGCATCTACGAAATG	719
Db	729	AAAGACATTTGATAGAGATGGTTTACTTGCAGGTGG-AAAATGGCATCTACGAAATG	788
Qy	720	CGACTGGAACATAGTGTGGATATAGATGTGG-ATATTTGATTTGGCTATTGCGAGCAA	777
Db	789	CGACTGGAACATAGTGTGGATATAGATGTGG-ATATTTGATTTGGCTATTGCGAGCAA	847
Qy	778	AGAGTATTAAAG-ATATGGCTATTTTGG	803
Db	848	AGAGTATTAAAGTATGCTATTTTGG	874

Search completed: May 11, 2004, 11:45:54
Job time : 3795.33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw mode1

Run on: May 11, 2004, 11:46:35 ; Search time 604.697 Seconds
(without alignments)
9776.195 Million cell updates/sec

Title: US-09-930-440B-3

Perfect score: 1305

Sequence: 1 atggactcgttgaggaagg.....ataattcatgccaaaataag 1305

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 6

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/PCTUS_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB3.seq:*
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
14:	/cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16:	/cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB3.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1303.4	99.9	1305	9	US-09-984-205-1
2	1303.4	99.9	1305	9	US-09-930-440B-3
3	1272.4	97.5	1376	13	US-10-342-887-1801
4	1272.4	97.5	1376	13	US-10-172-118-1801
5	1272.4	97.5	1770	15	US-10-037-270-360
6	1272.4	97.5	1770	16	US-10-117-722-360
7	1272.4	97.5	1800	9	US-09-925-301-534
8	686.8	52.6	738	16	US-10-012-697-1444
9	647.6	49.6	789	16	US-10-012-697-1369
10	627.4	48.1	745	16	US-10-012-697-394
11	620.2	47.5	748	16	US-10-012-697-569
12	617.4	47.3	708	16	US-10-012-697-109
13	562	43.1	686	16	US-10-012-697-1128
14	555.6	42.6	639	9	US-09-867-550-1193

15	400.6	30.7	408	10	US-09-918-995-36321
16	306.8	23.5	509	9	US-09-783-590-293
17	217	16.6	263	9	US-09-960-352-9191
18	185.6	14.2	251	9	US-09-796-692-3842
19	185.6	14.2	251	15	US-10-040-862-3842
20	185.6	14.2	251	16	US-10-057-475B-3842
21	185.6	14.2	251	16	US-10-154-884B-3842
22	145.8	11.2	222	11	US-09-864-408A-3237
23	145.4	11.1	205	9	US-09-960-352-12271
24	138.2	10.6	478	9	US-09-864-761-2461
25	134	10.3	134	9	US-09-864-761-19192
26	96.4	7.4	190	10	US-09-822-846-374
27	79.2	6.1	95	9	US-09-864-761-11953
28	77.4	5.9	397	9	US-09-864-761-15439
29	60	4.6	60	10	US-09-903-975-5141
30	50.6	3.9	1830121	15	US-10-329-960-1
31	50.6	3.9	1830121	16	US-10-329-960-1
32	50.2	3.8	495	13	US-10-335-977-2927
33	50.2	3.8	1296	15	US-10-156-761-3361
34	50.2	3.8	9025608	15	US-10-156-761-1
35	45.8	3.5	7462	15	US-10-311-455-1757
36	45.8	3.5	580073	15	US-10-205-220-1
37	44.6	3.4	40862	15	US-10-311-455-2046
38	42.4	3.2	2430	15	US-10-156-761-2531
39	42.4	3.2	9025608	15	US-10-156-761-1
40	42	3.2	627	13	US-10-282-122A-25195
41	41.6	3.2	119211	13	US-10-672-787-40
42	41.6	3.2	3186778	13	US-10-027-632-174961
43	41.6	3.2	3186778	16	US-10-027-632-174961
44	41.4	3.2	5802	13	US-10-152-886-84
45	41	3.1	13249	15	US-10-311-455-89

ALIGNMENTS

RESULT 1
US-09-984-205-1
; Sequence 1, Application US/09984205
; Patent No. US20020317175A1
; GENERAL INFORMATION:
; APPLICANT: Coleman, Timothy A. et al.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PFS0SD1
; CURRENT APPLICATION NUMBER: US/09/984,205
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: PCT/US00/05325
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/516,143
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1302)
; NAME/KEY: SITE
; LOCATION: (397)..(399)
; OTHER INFORMATION: yac encodes Tyr or His
; NAME/KEY: SITE
; LOCATION: (406)..(408)
; OTHER INFORMATION: gkt encodes Gly or Val
; NAME/KEY: SITE
; LOCATION: (439)..(441)
; OTHER INFORMATION: yca encodes Ser or Pro
; NAME/KEY: SITE
; LOCATION: (505)..(507)
; OTHER INFORMATION: gkt encodes Gly or Val

US-09-984-205-1

Query Match 99.9%; Score 1303.4; DB 9; Length 1305;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGACTCGGTGGAGAGGGGGCCGACCTCCGCTCCAAACCCGCGGGGGCGACCGTCC 60
 DB 1 ATGGACTCGGTGGAGAGGGGGCCGACCTCCGCTCCAAACCCGCGGGGGCGACCGTCC 60

QY 61 CGGGCGGGCCCGCGAGCTGACGAGCAACTCTCGCGCGCGCGAGCGAGGTGTGGAG 120
 DB 61 CGGGCGGGCCCGCGAGCTGACGAGCAACTCTCGCGCGCGCGAGCGAGGTGTGGAG 120

QY 121 AAGCCCCCGACCTGCGAGCCCTTAATCTGCGCCCGGGAGGCGAGAAAGCAATCCCGCTG 180
 DB 121 AAGCCCCCGACCTGCGAGCCCTTAATCTGCGCCCGGGAGGCGAGAAAGCAATCCCGCTG 180

QY 181 AAGAACATTAAGCACTGCGCGGGGTCCCGCTCATTTGGCTGCTGCTGCGCGCCCTG 240
 DB 181 AAGAACATTAAGCACTGCGCGGGGTCCCGCTCATTTGCTGCTGCTGCGCGCCCTG 240

QY 241 GATTCAGGGCCCTCCAGAGTGTATGGGTTCGACAGACCAATGATGAATGAGATGTG 300
 DB 241 GATTCAGGGCCCTCCAGAGTGTATGGGTTCGACAGACCAATGATGAATGAGATGTG 300

QY 301 GCCAAACAATTTGGTGCAAGTTCATCGAAGACTTCTGAAGTTTCAAAAGACAGCTCT 360
 DB 301 GCCAAACAATTTGGTGCAAGTTCATCGAAGACTTCTGAAGTTTCAAAAGACAGCTCT 360

QY 361 ACCTCACTAGATGCCATCATAGAATTTCTTAATTATATATGAATGAGGKTGACATTTGAGGA 420
 DB 361 ACCTCACTAGATGCCATCATAGAATTTCTTAATTATATATGAATGAGGKTGACATTTGAGGA 420

QY 421 AATATCAAGCTCTTCYCATGTTTACATCTATCTGTTTCAAAAGAGTTCGAGAAATG 480
 DB 421 AATATCAAGCTCTTCYCATGTTTACATCTATCTGTTTCAAAAGAGTTCGAGAAATG 480

QY 481 ACTCGAAGAGGATATGATTCCTGCTGTTGAGAGCCCATCAGTTTCGATGG 540
 DB 481 ACTCGAAGAGGATATGATTCCTGCTGTTGAGAGCCCATCAGTTTCGATGG 540

QY 541 AGTGAATTCGAAAGGAGTTCGTGAAGTGCAGAACCTCTGAAATTAATCCAGCTAAA 600
 DB 541 AGTGAATTCGAAAGGAGTTCGTGAAGTGCAGAACCTCTGAAATTAATCCAGCTAAA 600

QY 601 CGGCTCTCGACAGACTGGATGGAGAAATTAATGAATGGCTCATTTTATTTGCT 660
 DB 601 CGGCTCTCGACAGACTGGATGGAGAAATTAATGAATGGCTCATTTTATTTGCT 660

QY 661 AAAAGACATTTGATAGATGGTTCATTCGAGGTGGAAATGCAATACGAAATGC 720
 DB 661 AAAAGACATTTGATAGATGGTTCATTCGAGGTGGAAATGCAATACGAAATGC 720

QY 721 GAGCTGGAACATGATGTGGATATAGATGATGATTTGATGGCTTATTCGAGCAAGA 780
 DB 721 GAGCTGGAACATGATGTGGATATAGATGATGATTTGATGGCTTATTCGAGCAAGA 780

QY 781 GTATTAGATATGCTATTTTGGCAAGAGAGCTTAAGAAATTAATCTTTGTTTGC 840
 DB 781 GTATTAGATATGCTATTTTGGCAAGAGAGCTTAAGAAATTAATCTTTGTTTGC 840

QY 841 AATATTGATGATGCTCACCACATGGCCACATTTATGATCAGAGACCAAAAGAAATA 900
 DB 841 AATATTGATGATGCTCACCACATGGCCACATTTATGATCAGAGACCAAAAGAAATA 900

QY 901 ATATCTTATGATGTAAGAAGTCTATTGGGATAAGTTTATTAAGAAGTGGATTGAG 960
 DB 901 ATATCTTATGATGTAAGAAGTCTATTGGGATAAGTTTATTAAGAAGTGGATTGAG 960

QY 961 GTGAGGCTAATCTCAGAAAGGGCCGCTTCAAGCAGAGCGCTGCTCTTTAAACTGGAT 1020
 DB 961 GTGAGGCTAATCTCAGAAAGGGCCGCTTCAAGCAGAGCGCTGCTCTTTAAACTGGAT 1020

RESULT 2

US-09-930-440B-3
 ; Sequence 3, Application US/09930440B
 ; Patent No. US20020142386A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Betenbaugh et al.
 ; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
 ; FILE REFERENCE: PF509P2
 ; CURRENT APPLICATION NUMBER: US/09/930,440B
 ; PRIORITY FILING DATE: 2001-08-16
 ; PRIOR APPLICATION NUMBER: 60/227,579
 ; PRIORITY FILING DATE: 2000-08-25
 ; PRIOR APPLICATION NUMBER: 09/516,793
 ; PRIORITY FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: 60/169,624
 ; PRIORITY FILING DATE: 1999-12-08
 ; PRIOR APPLICATION NUMBER: 60/122,582
 ; PRIORITY FILING DATE: 1999-03-02
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 1305
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1305)
 ; NAME/KEY: misc feature
 ; LOCATION: (397)..(399)
 ; OTHER INFORMATION: The 'yat' at location 397..399 encodes amino acid His, or Tyr.
 ; NAME/KEY: misc feature
 ; LOCATION: (406)..(408)
 ; OTHER INFORMATION: The 'gkt' at location 406..408 encodes amino acid Gly, or Val.
 ; NAME/KEY: misc feature
 ; LOCATION: (439)..(441)
 ; OTHER INFORMATION: The 'yca' at location 439..441 encodes amino acid Pro, or Ser.
 ; NAME/KEY: misc feature
 ; LOCATION: (505)..(507)
 ; OTHER INFORMATION: The 'gkt' at location 505..507 encodes amino acid Gly, or Val.
 ; US-09-930-440B-3

Query Match 99.9%; Score 1303.4; DB 9; Length 1305;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGACTCGGTGGAGAGGGGGCCGACCTCCGCTCCAAACCCGCGGGGGCGACCGTCC 60
 DB 1 ATGGACTCGGTGGAGAGGGGGCCGACCTCCGCTCCAAACCCGCGGGGGCGACCGTCC 60

QY 61 CGGGCGGGCCCGCGAGCTGACGAGCAACTCTCGCGCGCGCGAGCGAGGTGTGGAG 120
 DB 61 CGGGCGGGCCCGCGAGCTGACGAGCAACTCTCGCGCGCGCGAGCGAGGTGTGGAG 120

Db	798	AGTATTAGATATGCTATTTTGGCAAAGAAGCTTAGGAAAATAAAACTTTTGGTTTG	857
Qy	840	CAATATTGATGGATGCTCCACCAATGGCCACATTTATGTATCAGAGACCAAAAAGAANT	899
Db	858	CAATATTGATGGATGCTCCACCAATGGCCACATTTATGTATCAGAGACCAAAAAGAANT	917
Qy	900	AATATCTTATGATGTAAAGATGCTATTGGGATTAAGTTTATTAAGAAAACTGGTATTGA	959
Db	918	AATATCTTATGATGTAAAGATGCTATTGGGATTAAGTTTATTAAGAAAACTGGTATTGA	977
Qy	960	GGTCAGGCTAACTCAGAAAAGGGCTGTCTCAACGACAGCGCTGCTCTTTTAAAACCTGGA	1019
Db	978	GGTCAGGCTAACTCAGAAAAGGGCTGTCTCAACGACAGCGCTGCTCTTTTAAAACCTGGA	1037
Qy	1020	TTGCAAAATGGAAGTCAGTGATCAGACAAGCTAGCAGTCTTAGATGAATGGAGAAAAGA	1079
Db	1038	TTGCAAAATGGAAGTCAGTGATCAGACAAGCTAGCAGTCTTAGATGAATGGAGAAAAGA	1097
Qy	1080	AATGGGCGCTGTGCTGGAAAAGATGGCAATCTTTGGAAATGAAGTGTCTGATGAAGATG	1139
Db	1098	AATGGGCGCTGTGCTGGAAAAGATGGCAATCTTTGGAAATGAAGTGTCTGATGAAGATG	1157
Qy	1140	CTTGAACAGAGTGGGCGCTAAGTCGGCGCTCCGTGATGCGCTGTCTTACGCCCGAAGGC	1199
Db	1158	CTTGAACAGAGTGGGCGCTAAGTCGGCGCTCCGTGATGCGCTGTCTTACGCCCGAAGGC	1217
Qy	1200	TGTTGGATACATTTTSCAAATGTAATGTGGCGCTGGTGCCATCCGAGAAATTTGCAGAGCA	1259
Db	1218	TGTTGGATACATTTTSCAAATGTAATGTGGCGCTGGTGCCATCCGAGAAATTTGCAGAGCA	1277
Qy	1260	CATTGGCTACTAATGGAAAAGTTAATAATTCATGCGCCAAAATAG	1305
Db	1278	CATTGGCTACTAATGGAAAAGTTAATAATTCATGCGCCAAAATAG	1323

RESULT 5

[illegible]

```
QY 960 GGTGAGGCTAATCTCAGAAAGGCGCTGTTCCAGACGACGCTGCTCTTTTAAACTGGA 1019
Db 1064 GGTGAGGCTAATCTCAGAAAGGCGCTGTTCCAGACGACGCTGCTCTTTTAAACTGGA 1123
QY 1020 TTGCAAAATGGAAGTCAAGTGTATCAGACAGCTAGCACTGTAGATGAATGGAGAAAGA 1079
Db 1124 TTGCAAAATGGAAGTCAAGTGTATCAGACAGCTAGCACTGTAGATGAATGGAGAAAGA 1183
QY 1080 AATGGGCTGTGCTGGAAGAGTGGCATATCTTGGAAATGAAGTGTCTGTGATGAAGAGTG 1139
Db 1184 AATGGGCTGTGCTGGAAGAGTGGCATATCTTGGAAATGAAGTGTCTGTGATGAAGAGTG 1243
QY 1140 CTTGAAGAGTGGGCTAAGTGGGCTCTCTGCTGATGCTGTCTCTACGCCAGAGGC 1199
Db 1244 CTTGAAGAGTGGGCTAAGTGGGCTCTCTGCTGATGCTGTCTCTACGCCAGAGGC 1303
QY 1200 TGTGGATCATTTGCAATGAATGTAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGCAGAGCA 1259
Db 1304 TGTGGATCATTTGCAATGAATGTAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGCAGAGCA 1363
QY 1260 CATTGGCTACTAATGGAAAGGTTAATTAATTCATGCCAAAATAG 1305
Db 1364 CATTGGCTACTAATGGAAAGGTTAATTAATTCATGCCAAAATAG 1409

RESULT 6
US-10-117-722-363
; Sequence 360, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
; FILE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2BCIP
; CURRENT APPLICATION NUMBER: US/10/117,722
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,322
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pc_FL_genes Version 1.0
; SEQ ID NO 360
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (105)..(1409)
; US-10-117-722-363

Query Match 97.5%; Score 1272.4; DB 16; Length 1770;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;

QY 1 ATGGAATCGGTGGAGAGGGGGCGCCACCTCCGTCCTCAACCGCGGGGGCGACCGTCC 60
Db 105 ATGGAATCGGTGGAGAGGGGGCGCCACCTCCGTCCTCAACCGCGGGGGCGACCGTCC 164
QY 61 CGGGGCGGGCGCGAGAGCTGCAGGCACTCTCGCGCGCGCCAGGCGCGAGGTGTGAG 120
Db 165 CGGGGCGGGCGCGAGAGCTGCAGGCACTCTCGCGCGCGCCAGGCGCGAGGTGTGAG 224
QY 121 AAGCCCCCGCACCTGGCAGCCCTAATTTCTGSCCGGGGAGCGACGAAGGCTCCCGCTG 180
Db 225 AAGCCCCCGCACCTGGCAGCCCTAATTTCTGSCCGGGGAGCGACGAAGGCTCCCGCTG 284
```

```
QY 181 AAGAACATTAAGCACTGGCGGGGTCCTCCCTCATTTGGCTGGGTCCTGGGTCCTGGGTCCTG 240
Db 285 AAGAACATTAAGCACTGGCGGGGTCCTCCCTCATTTGGCTGGGTCCTGGGTCCTGGGTCCTG 344
QY 241 GATTTCAGGGGCTCTCCAGAGTGTATGGGTTTCCAGACAGCCATGATGAATTAAGAAATGTG 300
Db 345 GATTTCAGGGGCTCTCCAGAGTGTATGGGTTTCCAGACAGCCATGATGAATTAAGAAATGTG 404
QY 301 GCCAAACAAATTTGGTGCACAGTTTCATCGAAGAGTTCTGAAATTTCAAAAGACAGCTCT 360
Db 405 GCCAAACAAATTTGGTGCACAGTTTCATCGAAGAGTTCTGAAATTTCAAAAGACAGCTCT 464
QY 361 ACCTGCTAGTGCATCATAGAAATTTCTTAATTAATTAATGAGGTTGACATTTAGAA 420
Db 465 ACCTGCTAGTGCATCATAGAAATTTCTTAATTAATTAATGAGGTTGACATTTAGAA 524
QY 421 AATATTCAAGCTACTCTCTCATGTTTACATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
Db 525 AATATTCAAGCTACTCTCTCATGTTTACATCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 584
QY 481 ATTGGAAGAAAGGATATGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
Db 585 ATTGGAAGAAAGGATATGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 644
QY 541 ACTGGAATTCAGAAAGGAGTTCTGTAAGTGAACGAACTCTGTAATTAATTAATTAATTAAT 600
Db 645 ACTGGAATTCAGAAAGGAGTTCTGTAAGTGAACGAACTCTGTAATTAATTAATTAATTAAT 704
QY 601 CGGCTCTCTGCGCAAGACTGGGATGGAGAAATTTATATGAATGAATGGCTCATTTTATTTGCT 660
Db 705 CGGCTCTCTGCGCAAGACTGGGATGGAGAAATTTATATGAATGAATGGCTCATTTTATTTGCT 764
QY 661 AAGAGCANTTTGATAGAGAGTGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 719
Db 765 AAGAGCANTTTGATAGAGAGTGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 824
QY 720 CGAGCTGGAACATAGTGTGATATAGTGTGATATGATGATGATGATGATGATGATGATGATG 779
Db 825 CGAGCT - GAACATAGTGTGATATAGTGTGATATGATGATGATGATGATGATGATGATGATG 883
QY 780 AGTATTAAAGATATGCTATTTTGGCAAGAGAGAGCTTAAGGAAATTAAGAAATTAAGAAAT 839
Db 884 AGTATTAAAGATATGCTATTTTGGCAAGAGAGAGCTTAAGGAAATTAAGAAATTAAGAAAT 943
QY 840 CAATATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 899
Db 944 CAATATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1003
QY 900 AATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 959
Db 1004 AATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1063
QY 960 GGTGAGGCTAATCTCAGAAAGGCGCTGTTCCAGACGACGCTGCTCTCTTTAAACTGGA 1019
Db 1064 GGTGAGGCTAATCTCAGAAAGGCGCTGTTCCAGACGACGCTGCTCTCTTTAAACTGGA 1123
QY 1020 TTGCAAAATGGAAGTCAAGTGTATCAGACAGCTAGCACTGTAGATGAATGGAGAAAGA 1079
Db 1124 TTGCAAAATGGAAGTCAAGTGTATCAGACAGCTAGCACTGTAGATGAATGGAGAAAGA 1183
QY 1080 AATGGGCTGTGCTGGAAGAGTGGCATATCTTGGAAATGAAGTGTCTGTGATGAAGAGTG 1139
Db 1184 AATGGGCTGTGCTGGAAGAGTGGCATATCTTGGAAATGAAGTGTCTGTGATGAAGAGTG 1243
QY 1140 CTTGAAGAGTGGGCTAAGTGGGCTCTCTGCTGATGCTGTCTCTACGCCAGAGGC 1199
Db 1244 CTTGAAGAGTGGGCTAAGTGGGCTCTCTGCTGATGCTGTCTCTACGCCAGAGGC 1303
QY 1200 TGTGGATCATTTGCAATGAATGTAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGCAGAGCA 1259
Db 1304 TGTGGATCATTTGCAATGAATGTAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGCAGAGCA 1363
QY 1260 CATTGGCTACTAATGGAAAGGTTAATTAATTCATGCCAAAATAG 1305
```

Db 1364 CATTGGCTACTAATGNAAGGTTAATAATTCATGCCAAAATAG 1409

RESULT 7

US-09-925-301-534
 ; Sequence 534, Application US/09925301
 ; Patent No. US20020052308A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA106
 ; CURRENT APPLICATION NUMBER: US/09/925,301
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05982
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 1694
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 534
 ; LENGTH: 1800
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-925-301-534

Query Match 97.5%; Score 1272.4; DB 9; Length 1800;
 Best Local Similarity 99.2%; Pred. No. 0;
 Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;

QY	1	ATGACCTCGGTGGAGAGGGGGCGCCACCTCGCTCTCCAAACCGCGGGGGCGACCGTCC	60
DB	120	ATGACCTCGGTGGAGAGGGGGCGCCACCTCGCTCTCCAAACCGCGGGGGCGACCGTCC	179
QY	61	CGGGGGCGGGCGGCAAGCTGCGCGGCAACTCTCGCGGGGGCGGCGGCGGAGGTGGAG	120
DB	180	CGGGGGCGGGCGGCGGCAAGCTGCGCGGCAACTCTCGCGGGGGCGGCGGCGGAGGTGGAG	239
QY	121	AGCCCCCGCCACCTGCGACCGCTTAATCTGCGCGGGGGCGGCGGCGGAGGTGGAG	180
DB	240	AGCCCCCGCCACCTGCGACCGCTTAATCTGCGCGGGGGCGGCGGCGGAGGTGGAG	299
QY	181	AGAAACATTAAGCACCTGGCGGGGCTCCGCTCAATGGCTGGGCTCTCGCTGGGGCGCTG	240
DB	300	AGAAACATTAAGCACCTGGCGGGGCTCCGCTCAATGGCTGGGCTCTCGCTGGGGCGCTG	359
QY	241	GATTCAGGGGCTTCAGAGTGTATGGGTTTCAGACAGCAATGATGAATGGAGATGTG	300
DB	360	GATTCAGGGGCTTCAGAGTGTATGGGTTTCAGACAGCAATGATGAATGGAGATGTG	419
QY	301	GCCAAACATTTGGTGCACAAAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTCT	360
DB	420	GCCAAACATTTGGTGCACAAAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTCT	479
QY	361	ACCTCACTAGATGCCATCATAGAAATTTCTTAATATATATGAGGKTCACATTTGAGGA	420
DB	480	ACCTCACTAGATGCCATCATAGAAATTTCTTAATATATATGAGGKTCACATTTGAGGA	539
QY	421	AATATTCAGCTACTTCTTCATGTTTACATTCATGATCTTCAAAAAGTTTCAGAAATG	480
DB	540	AATATTCAGCTACTTCTTCATGTTTACATTCATGATCTTCAAAAAGTTTCAGAAATG	599
QY	481	ATTGAGAGAGAGGATCATGATCTCTGTTTCTGTTGTCAGAGCGCATCAGTTTCATGAG	540
DB	600	ATTGAGAGAGAGGATCATGATCTCTGTTTCTGTTGTCAGAGCGCATCAGTTTCATGAG	659
QY	541	AGTGAATTCAGAAAGGAGTTCGTGAAGTGAACCGACCTCTCAATTTAAATCCAGCTAA	600
DB	660	AGTGAATTCAGAAAGGAGTTCGTGAAGTGAACCGACCTCTCAATTTAAATCCAGCTAA	719
QY	601	CGGCTCTGTCGACAGCTGGAGTGGAGATTTATATGAAAATGGCTCAATTTTATTTGCT	660
DB	720	CGGCTCTGTCGACAGCTGGAGTGGAGATTTATATGAAAATGGCTCAATTTTATTTGCT	779

QY	661	AAAAGACATTTGATAGACATGGGTTTACTTGCAGGCTGG-AAAATGGCATACTACGAATG	719
DB	780	AAAAGACATTTGATAGACATGGGTTTACTTGCAGGCTGGAAAANTGSCATACTACGAATG	839
QY	720	CGAGCTGGACATAGTGTGGATATAGCTGTGGATATTTGATTTGGCTTATTCAGAGCAAG	779
DB	840	CGAGCT-GAACATAGTGTGGATATAGCTGTGGATATTTGATTTGGCTTATTCAGAGCAAG	898
QY	780	AGTATTAGATATGCTATTTTGGCAAGAGAGCTTAAAGAAATAAAACATTTTGGTTG	839
DB	899	AGTATTAGATATGCTATTTTGGCAAGAGAGCTTAAAGAAATAAAACATTTTGGTTG	958
QY	840	CAATATTGATGATGTCTCACCATTGGCCACATTTATGATATCAGGAGACCAAAAAGAA	899
DB	959	CAATATTGATGATGTCTCACCATTGGCCACATTTATGATATCAGGAGACCAAAAAGAA	1018
QY	900	AATATCTTATGATGTAAAGAGTGTATTTGGGATAAGTTTATTAAGAAAGTGGTATGA	959
DB	1019	AATATCTTATGATGTAAAGAGTGTATTTGGGATAAGTTTATTAAGAAAGTGGTATGA	1078
QY	960	GGTGAGCTAATCTCAGAAAGGGGCTGTTCAAAGCAGACGCTCTCTTTTAAACTGGA	1019
DB	1079	GGTGAGCTAATCTCAGAAAGGGGCTGTTCAAAGCAGACGCTCTCTTTTAAACTGGA	1138
QY	1020	TTGCAAAATGGAAGTCAAGTGTATCAGACCAAGCTAGCAATGATGATGAGAGAAAGA	1079
DB	1139	TTGCAAAATGGAAGTCAAGTGTATCAGACCAAGCTAGCAATGATGATGAGAGAAAGA	1198
QY	1080	AATGGGCTCTGCTGGAAAGAGTGGCATATCTTGGAAATGAAGTGTCTGATGAGAGTG	1139
DB	1199	AATGGGCTCTGCTGGAAAGAGTGGCATATCTTGGAAATGAAGTGTCTGATGAGAGTG	1258
QY	1140	CTTGAAGAGAGTGGGCTTAAGTGGCGCTCCGCTGATGCTCTTCTACGCCAGAGAGGC	1199
DB	1259	CTTGAAGAGAGTGGGCTTAAGTGGCGCTCCGCTGATGCTCTTCTACGCCAGAGAGGC	1318
QY	1200	TGTTGGATACATTTGCAAAATGTAATGCTGGCGCTGGTGGCCATCCGAGAAATTCAGAG	1259
DB	1319	TGTTGGATACATTTGCAAAATGTAATGCTGGCGCTGGTGGCCATCCGAGAAATTCAGAG	1378
QY	1260	CATTGGCTACTAATGCAAAAGTTTATATTAATTCATGCAAAATAG 1305	
DB	1379	CATTGGCTACTAATGCAAAAGTTTATATTAATTCATGCAAAATAG 1424	

RESULT 8

US-10-012-697-1444
 ; Sequence 1444, Application US/10012697
 ; Publication No. US20030215803A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Escobedo, Jaime
 ; APPLICANT: Garcia, Pablo Dominguez
 ; APPLICANT: Kassam, Altaf
 ; APPLICANT: Lamson, George
 ; APPLICANT: Scott, Beth
 ; APPLICANT: Drmanac, Radoje
 ; APPLICANT: Crkvenjakov, Radomir
 ; APPLICANT: Dickson, Mark
 ; APPLICANT: Drmanac, Snezana
 ; APPLICANT: Labat, Ivan
 ; APPLICANT: Leshkowitz, Dena
 ; APPLICANT: Kita, David
 ; APPLICANT: Garcia, Veronica
 ; APPLICANT: Jones, Lee William
 ; APPLICANT: Stache-Crain, Birgit
 ; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS
 ; FILE REFERENCE: 2300-16252
 ; CURRENT APPLICATION NUMBER: US/10/012,697
 ; PRIOR FILING DATE: 2003-02-21
 ; PRIOR APPLICATION NUMBER: 60/254,648
 ; PRIOR FILING DATE: 2000-12-07

;; PRIOR APPLICATION NUMBER: 60/275,668
;; PRIOR FILING DATE: 2001-03-13
;; NUMBER OF SEQ ID NOS: 1568
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 1444
;; LENGTH: 738
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-012-697-1444

Query Match 52.6%; Score 686.8; DB 16; Length 738;
Best Local Similarity 97.3%; Pred. No. 1.2e-182;
Matches 716; Conservative 3; Mismatches 15; Indels 2; Gaps 2;

```
QY 405 GGTGACATGTTAGGAATAATTCACCTACTCTCTCATCTTTACATCTTACCTCTACTGATCTTCA 464
DB 3 GGTGACATGTTAGGAATAATTCACCTACTCTCTCATCTTTACATCTTACCTCTACTGATCTTCA 62
QY 465 AAAAGTTGCAGAAATGANTTCGAGAGAGAGATATGATTTCTGKTTTCTCTGTTGTGAGAGC 524
DB 63 AAAAGTTGCAGAAATGANTTCGAGAGAGAGATATGATTTCTGTTTCTCTGTTGTGAGAGC 122
QY 525 CCATCAGTTTCATGAGAGTGAATTCAGAAAGAGAGTCTGTGAAGTACCGAACCTCTGAA 584
DB 123 CCATCAGTTTCATGAGAGTGAATTCAGAAAGAGTCTGTGAAGTACCGAACCTCTGAA 182
QY 585 TTAAATCCAGCTAAACGGCTCGTCGACAGACTGGGATGGAGATATATGAAATGG 644
DB 183 TTAAATCCAGCTAAACGGCTCGTCGACAGACTGGGATGGAGATATATGAAATGG 242
QY 645 CTCATTTTATTGTTGCTAAAGACATTTGATGAGATGGGTACTTTCGAGGGTGG-AAAT 703
DB 243 CTCATTTTATTGTTGCTAAAGACATTTGATGAGATGGGTACTTTCGAGGGTGGAAAT 302
QY 704 GGCATACCTCGAATGCGAGCTGGAACATAGTGTGATATGATGATGATGATGATGATGATGATG 763
DB 303 GGCATACCTCGAATGCGAGCTGGAACATAGTGTGATATGATGATGATGATGATGATGATGATG 361
QY 764 CTATTGCAGAGCAAGAGTATTAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGAA 823
DB 362 CTATTGCAGAGCAAGAGTATTAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGAA 421
QY 824 TAAACTTTTGGTTTGCATATATGATGATGATGCTCACCAATGGCCACATTTATGATATCAG 883
DB 422 TAAACTTTTGGTTTGCATATATGATGATGATGCTCACCAATGGCCACATTTATGATATCAG 481
QY 884 GAGACCAAAAGAAATATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 943
DB 482 GAGACCAAAAGAAATATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 541
QY 944 AGAAAGTGGTATTGAGGTGAGGCTAATCTCAGAAAGGGCTGTTCAAAGCAGAGCTGT 1003
DB 542 AGAAAGTGGTATTGAGGTGAGGCTAATCTCAGAAAGGGCTGTTCAAAGCAGAGCTGT 601
QY 1004 CTTCTTTAAACTGGATGCAAAATGGAAGTCTGTATCAGCAAGTACAGTACAGTTGAG 1063
DB 602 CTTCTTTAAACTGGATGCAAAATGGAAGTCTGTATCAGCAAGTACAGTACAGTTGAG 661
QY 1064 ATGAATGGAGAAAGAAATGGGCTGTGCTGGAAGAGTGGCATATCTTGGAAATGAG 1123
DB 662 ATGAATGGAGAAAGAAATGGGCTGTGCTGGAAGAGTGGCATATCTTGGAAATGAG 721
QY 1124 TGTCTGATGAGAGTG 1139
DB 722 TACTTGTAGGAATG 737
```

RESULT 9
US-10-012-697-1369
; Sequence 1369, Application US/10012697
; Publication No. US20030215803A1
; GENERAL INFORMATION:
; APPLICANT: Escobedo, Jaime

;; APPLICANT: Garcia, Pablo Dominguez
;; APPLICANT: Kassam, Altaf
;; APPLICANT: Lamson, George
;; APPLICANT: Scott, Beth
;; APPLICANT: Drmanac, Radoje
;; APPLICANT: Crkvenjakov, Radomir
;; APPLICANT: Dickson, Mark
;; APPLICANT: Drmanac, Snezana
;; APPLICANT: Labat, Ivan
;; APPLICANT: Leshkowitz, Dena
;; APPLICANT: Kita, David
;; APPLICANT: Garcia, Veronica
;; APPLICANT: Jones, Lee William
;; APPLICANT: Stache-Crain, Birgit
;; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS
;; FILE REFERENCE: 2300-16252
;; CURRENT APPLICATION NUMBER: US/10/012,697
;; CURRENT FILING DATE: 2003-01-21
;; PRIOR APPLICATION NUMBER: 60/254,648
;; PRIOR FILING DATE: 2000-12-07
;; PRIOR APPLICATION NUMBER: 60/275,668
;; PRIOR FILING DATE: 2001-03-13
;; NUMBER OF SEQ ID NOS: 1568
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 1369
;; LENGTH: 789
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 2, 9, 11, 23, 33, 756, 771
;; OTHER INFORMATION: n = A,T,C or G
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 2, 9, 11, 23, 33, 756, 771
;; OTHER INFORMATION: n = A,T,C or G
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 2, 9, 11, 23, 33, 756, 771
;; OTHER INFORMATION: n = A,T,C or G
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 2, 9, 11, 23, 33, 756, 771
;; OTHER INFORMATION: n = A,T,C or G
US-10-012-697-1369

Query Match 49.6%; Score 647.6; DB 16; Length 789;
Best Local Similarity 94.8%; Pred. No. 1.4e-171;
Matches 729; Conservative 5; Mismatches 22; Indels 13; Gaps 6;

```
QY 150 GGCCCGGGAGGAGCAGCAAGGSCATCCCTCGAAGAACATTAAGCACCTGGCGGGGTC 209
DB 14 GGCCCGGGAGGAGCAGCAAGGSCATCCCGGGAGAACATTAAGCACCTGGCGGGGTC 73
QY 210 GCTCATTTGGTGGTCTCTGCTGCGGCCCTTGATTCAGGGGCTTCCAGAGTGTATGGT 269
DB 74 GCTCATTTGGTGGTCTCTGCTGCGGCCCTTGATTCAGGGGCTTCCAGAGTGTATGGT 133
QY 270 TTGACAGACCATGATGAATTTGAGATGTGGCCAAACAATTTGGTGGACAAAGTTCTAT 329
DB 134 TTGACAGACCATGATGAATTTGAGATGTGGCCAAACAATTTGGTGGACAAAGTTCTAT 193
QY 330 AAGAAGTCTGAAGTTTCBAAGACAGCTCTACTCTACTAGATGCCATCATAGATTTCT 389
DB 194 AAGAAGTCTGAAGTTTCBAAGACAGCTCTACTCTACTAGATGCCATCATAGATTTCT 253
QY 390 TAATTATYATAATCAGGKTGACATTTAGGAATATTTCAAGTCTACTTCTTCATCTTACA 449
```


254	Db	TATATTATCATTAATGAGGTTGACATTGTAGGAANAATTCAAGCTACTCTTCCTCATGTTTACA	313
450	Qy	TCCTACTGATCTTCAAAAAGTTGCAGAAATGATTCGAGAAGAAGGATATGATCTCTGKTTT	509
314	Db	TCCTACTGATCTTCAAAAAGTTGCAGAAATGATTCGAGNAGNAGGATATGATCTCTGTTTT	373
510	Qy	CTCTGTTGTGAGAGCCCATCAGTTTCGATGGAGTGAAATTCAGAAAGAGTTTCGTGAAGT	569
374	Db	CTCTGTTGTGAGAGCCCATCAGTTTCGATGGAGTGAAATTCAGAAAGAGTTTCGTGAAGT	433
570	Qy	GACCGAACCCTCGAATTTAAATCCAGCTTAAACGGGCTCGTCGACAAGACTCGGATGGAGA	629
434	Db	GACCGAACCCTCGAATTTAAATCCAGCTTAAACGGGCTCGTCGACAAGACTCGGATGGAGA	493
630	Qy	ATTATATGAAATCGGCTCAATTTTATTTTGCTTAAAGAATTGTGATAGAGATGGGTACTT	689
494	Db	ATTATATGAAATCGGCTCAATTTTATTTTGCTTAAAGAATTGTGATAGAGATGGGTACTT	553
690	Qy	GCAGGGTGG-AAAATGGCATACTACGAAATCGAGCTCGAAACATAGTGTGGATATAGATG	748
554	Db	GCAGGGTGGAAATATGGCATACTACGAAATCGAGCT- GAACATAGTGTGGATATAGATG	612
749	Qy	TGGATATTGATTGGCTTATTGCAGAGCAAGAGTATTAAAGTATGGCTATTTTGSCAAAG	808
613	Db	TGGATATTGATTGGCTTATTGCAGAGCAAGAGTATTAAAGTATGGCTATTTTGSCAAAG	672
809	Qy	AGAAGCTTAAGGAATAAAACTTTTGGTTTGCATATTATGATGGATGCTCTCCACCATGGCC	868
673	Db	AGAAGCTTAAGGA- -TAACTTTTGGTT- -GCATATTATGATGGATGCTCTCCACCATGGCC	728
869	Qy	ACATTTATGATTCAGGAGACCAAAAAGAAATAATATCTTATGATGTAAA	917
729	Db	A- - -TTTGTATCAGGAGACCA- - -AGATATNTCTTATGATGTAAA	779

RESULT 10

```

US-10-012-697-394
; Sequence 394, Application WS/10012697
; Publication No. US2003021583A1
GENERAL INFORMATION:
; APPLICANT: Escobedo, Jaime
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Scott, Beth
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS
; TITLE OF INVENTION: ISOLATED FROM HUMAN PROSTATE
; FILE REFERENCE: 2360-16252
; CURRENT APPLICATION NUMBER: US/10/012,697
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/254,648
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 60/275,668
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 1568
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 394
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2, 9, 11, 23, 33, 627, 652, 673, 727-731, 735

```



```
QY 310 TTTGGTGCACAGTTTCATCGAAGAGTTCTGGAAGTTTCAAAGAGACAGCTCTACCTCACCCTA 369
Db 124 TTGGTGCACAGTTTCATCGAAGAGTTCTGGAAGTTTCAAAGAGACAGCTCTACCTCACCCTA 183
QY 370 GATGCCATCATAGAAATTTCTTAATTATATATATGATGAGGKTGACATTTGAGGAATATTCNA 429
Db 184 GATGCCATCATAGAAATTTCTTAATTATATATATGATGAGGKTGACATTTGAGGAATATTCNA 243
QY 430 GCTACTTCTCATGTTTACATCTTACTACTGATCTTTCAAAAAGTTTGCAAAAATGATTCGAGAA 489
Db 244 GCTACTTCTCATGTTTACATCTTACTACTGATCTTTCAAAAAGTTTGCAAAAATGATTCGAGAA 303
QY 490 GAAGCATATGATTTCTGKTTTCTCTGTTGGAAGCCCATCAGTTTCGATGAGTGAATTT 549
Db 304 GAAGCATATGATTTCTGKTTTCTCTGTTGGAAGCCCATCAGTTTCGATGAGTGAATTT 363
QY 550 CAGAAAAGAGTTCGTGAAGTGACCGAAGCTCTGAAATTTAAATTCAGCTAAACGGGCTCGT 609
Db 364 CAGAAAAGAGTTCGTGAAGTGACCGAAGCTCTGAAATTTAAATTCAGCTAAACGGGCTCGT 423
QY 610 CGACAGACTGGGATGGGAATATATGAAAATGCTCATTTATTTTCTTAAAGACAT 669
Db 424 CGACAGACTGGGATGGGAATATATGAAAATGCTCATTTATTTTCTTAAAGACAT 483
QY 670 TTGATAGAGAT-GGGTTACTTTGCGAGGTGGAAAATGGCATACTAGAAAATGGAGCTCGA 728
Db 484 TTGATAGAGATGGGTTACTTTGCGAGGTGGAAAATGGCATACTAGAAAATGGAGCTCGA 542
QY 729 ACATAGTGTGATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 788
Db 543 ACATAGTGTGATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 602
QY 789 ATATGGCTATTTTGGCAAGAGAGAG 813
Db 503 NIGGCTN-TTTTGCRAANAGCTTAAG 627
```

RESULT 14

```
US-09-867-550-1193
; Sequence 1193, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehrahan, Foad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; PRIOR FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1193
; LENGTH: 639
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(54)
; OTHER INFORMATION: Wherein any n is one of a or t or c or g
US-09-867-550-1193
```

```
Query Match 42.6%; Score 555.6; DB 9; Length 639;
Best Local Similarity 94.8%; Pred. No. 1.1e-145;
Matches 580; Conservative 4; Mismatches 22; Indels 6; Gaps 1;
QY 1 ATGACATCGGTGGAGAGGGGGCCGACCTCGCTCTCCACCCCGGGGGCGACCGTCC 60
Db 34 ATGACATCGGTGGAGAGGGGGCCGACCTCGCTCTCCACCCCGGGGGCGACCGTCC 93
```

```
QY 61 CGGGCGCGGCGCGCGAGCTGAGCGCAACTCTCCGCGCGCGCGCGCGCGAGTGTGGAG 120
Db 94 CGGGCGCGGCGCGCGAGCTGAGCGCG-----ATCTCGCGCGCGCGCGCGAGTGTGGAG 147
QY 121 AAGCCCCCGCAGCTGCGAGCCCTTAATTTCTGGCCCCGGGAGGCGAGCAAAAGGCATCCCCCTG 180
Db 148 AAGCCCNCGACCTGCGAGCCCTTAATTTCTGGCCCCGGGAGGCGAGCAAAAGGCATCCCCCTG 207
QY 181 AAGAACATTAAAGCACTGCGCGGGGTCCCGCTCAITGGTGGGTCTCTGGTGGGCGCTG 240
Db 208 AAGAACATTAAAGCACTGCGCGGGGTCCCGTTCATTTGGTGGGTCTCTGGTGGGCGCTG 267
QY 241 GATTCAGGGGCTTCCAGAGTGTATGGGTTCGACAGACCATTGATGAATTTGAGAAATG 300
Db 268 GATTCAGGGGCTTCCAGAGTGTATGGGTTCGACAGACCATTGATGAATTTGAGAAATG 327
QY 301 GCCAAACAATTTGGTGCAAGTTTCATCGAAGAGTTCGAGAGTTTCAAAAGACAGCTCT 360
Db 328 GCCAAACAATTTGGTGCAAGTTTCATCGAAGAGTTCGAGAGTTTCAAAAGACAGCTCT 387
QY 361 ACCTCAGTAGATGCCATCATAGATTTCTTAATTAATATATAGAGTTCAGAGTTCAGAG 420
Db 388 ACCTCAGTAGATGCCATCATAGATTTCTTAATTAATATATAGAGTTCAGAGTTCAGAG 447
QY 421 AATATTCAAGTACTTCTTCATGTTTACTCTGCTACTGATCTTCAAAAGTTCGAGAAATG 480
Db 448 AATATTCAAGTACTTCTTCATGTTTACTCTGCTACTGATCTTCAAAAGTTCGAGAAATG 507
QY 481 ATTCGAGAAGAGGATATGATTTCTGKTTTCTCTGTTGTGAGACGCCATCAGTTTCGATGG 540
Db 508 ATTCGAGAAGAGGATATGATTTCTGTTTCTCTGTTGTGAGACGCCATCAGTTTCGATGG 567
QY 541 AGTGAAATTCAGAAAGAGGAGTTCTGTAAGTGACCGAACCTCTGAATTTAAATCCAGCTAAA 600
Db 568 AGTGAAATTCAGAAAGAGGAGTTCTGTAAGTGACCGAACCTCTGAATTTAAATCCAGCTAAA 627
QY 601 CGGCTCTGTCGA 612
Db 628 CGGCTCTGTCGA 639
```

RESULT 15

```
US-09-918-995-36321
; Sequence 36321, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36321
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-36321
```

```
Query Match 30.7%; Score 400.6; DB 10; Length 408;
Best Local Similarity 99.0%; Pred. No. 3.9e-102;
Matches 403; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 871 ATTATCTATCAGAGACCAAAAGAAATATATCTTATCATGTATAAAGATGCTATTGGG 930
Db 1 ATTATCTATCAGAGACCAAAAGAAATATATCTTATCATGTATAAAGATGCTATTGGG 60
QY 931 ATAAGTTTATTAAGAAAGTGGTATTGAGGTGAGGTAAATCTCAGAAAAGGCGCTGTCA 990
```

Db 61 ATAGTTATTAAAGAAAGTGGTATTGAGGTGAGGCTAATCTCGAAAGGGCCTTTCA 120
Qy 991 AAGCAGACGCTGTCTTCTTTAAACTGGATTGCAAAATGGAAGTCAGTGTATCAGACAAG 1050
Db 121 AAGCAGACGCTGTCTTCTTTAAACTGGATTGCAAAATGGAAGTCAGTGTATCAGACAAG 180
Qy 1051 CTAGCAGTTGTAGTGAATGGAGAAAGAAATGGGCTGTGCTGGAAGAGTGGCATAT 1110
Db 181 CTAGCAGTTGTAGTGAATGGAGAAAGAAATGGGCTGTGCTGGAAGAGTGGCATAT 240
Qy 1111 CTTGGAATGAAAGTGTCTCATGAAGAGTCTTGAAGAGAGTGGGCTTAAGTGGCGCTCT 1170
Db 241 CTTGGAATGAAAGTGTCTCATGAAGAGTCTTGAAGAGAGTGGGCTTAAGTGGCGCTCT 300
Qy 1171 GCTGATGCCCTGTTCCTACGCCAGAGGCTGTGGATACATTTGCAAAATGTAATGGTGGC 1230
Db 301 GCTGATGCCCTGTTCCTACGCCAGAGGCTGTGGATACATTTGCAAAATGTAATGGTGGC 360
Qy 1231 CGTGGTGCATCCGAGAAATTGCAGAGCACATTTGCCTACTAATGGA 1277
Db 361 CGTGGTGCATCCGAGAAATTGCAGAGCACATTTGCCTACTAATGGA 407

Search completed: May 12, 2004, 12:37:19
Job time : 624.697 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:55:43 ; Search time 61.1848 Seconds
(without alignments)
2004.184 Million cell updates/sec

Title: US-09-930-440B-4
Perfect score: 2266
Sequence: 1 MSVEKGAATSVNPRGRPS.....REFAHICLLMEKVNNSQK 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 34
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_29Jan04:*
- 1: Geneseq1980s:*
 - 2: Geneseq1990s:*
 - 3: Geneseq2000s:*
 - 4: Geneseq2001s:*
 - 5: Geneseq2002s:*
 - 6: Geneseq2003as:*
 - 7: Geneseq2003bs:*
 - 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2258	99.6	434	3 AAY96100	Human CMP
2	2258	99.6	434	3 AAY90351	Human gly
3	2258	99.6	434	4 AAB84682	Nucleotid
4	2258	99.6	434	6 AAO26545	Human CMP
5	2188	96.6	434	3 AAB28673	Human car
6	2188	96.6	434	4 AAM39319	Human pol
7	2188	96.6	434	4 AAB35213	Human pro
8	2188	96.6	448	3 AAB43931	Human can
9	2181	96.2	434	3 AAB43165	Human ORF
10	2149	94.8	461	4 AAM41105	Human pol
11	1173	51.8	245	7 ADB82754	Human pro
12	945	41.7	202	5 AAB64227	Human ORF
13	419	18.5	144	4 AAU29738	Novel hum
14	339	15.0	74	5 AAB32676	Human ORF
15	241	10.6	44	4 AAM16058	Peptide #
16	241	10.6	44	4 ABB35050	Peptide #
17	241	10.6	44	4 ABB29866	Peptide #
18	241	10.6	44	4 AAB20462	Protein #
19	241	10.6	44	4 AAM68236	Human bon
20	241	10.6	44	4 AAM55865	Human bra
21	241	10.6	44	4 ABB49887	Human liv
22	241	10.6	44	4 AAM03785	Peptide #
23	228	10.1	188	6 AAM69485	Phototab
24	225.5	10.0	169	3 AAB18422	Amino aci
25	217	9.6	228	3 AAY50798	N. mening

26	210.5	9.3	178	6 ABP80888	Abp80888 N. gonorr
27	208.5	9.2	410	3 AAY68969	Aay68969 Cps2r whi
28	198.5	8.8	413	5 ABP226813	Abp226813 Streptoco
29	177.5	7.8	209	6 ABU33455	Abu33455 Protein e
30	169.5	7.5	221	3 AAY97209	Aay97209 Campyloba
31	169.5	7.5	221	6 ABJ18499	Abj18499 Campyloba
32	159	7.0	180	6 ADA35085	Ada35085 Acinetoba
33	154	6.8	1012	4 ABG18198	Abg18198 Novel hum
34	154	6.8	1012	4 ABG28888	Abg28888 Novel hum
35	145	6.4	536	6 ABJ18487	Abj18487 Campyloba
36	145	6.4	536	6 ABJ18490	Abj18490 Campyloba
37	138.5	6.1	498	4 ABG25738	Abg25738 Novel hum
38	137.5	6.1	165	2 AAW20503	Aaw20503 H. pylori
39	128	5.6	467	2 AAR21577	Aar21577 HCV CKS-N
40	128	5.6	467	2 AAR33644	Aar33644 HCV CKS-N
41	128	5.6	467	2 AAR33606	Aar33606 HCV CKS-N
42	128	5.6	467	2 AAR33586	Aar33586 HCV CKS-N
43	128	5.6	467	4 AAB69005	Aab69005 HCV recom
44	128	5.6	541	2 AAR21567	Aar21567 HCV CKS-N
45	128	5.6	541	2 AAR33634	Aar33634 HCV CKS-N

ALIGNMENTS

RESULT 1
AAY96100
ID AAY96100 standard; protein; 434 AA.
XX
AC AAY96100;
XX
DT 19-DEC-2000 (first entry)
XX
DE Human CMP-sialic acid synthetase.
XX
KW CMP-sialic acid synthetase; human; sialylation; glycoprotein;
KW plasminogen; transferrin; thyrotropin; Na+,K+-ATPase.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FT Misc-difference 133 /note= "encoded by YAR"
FT Misc-difference 136 /note= "encoded by GKT"
FT Misc-difference 147 /note= "encoded bu YCA"
FT Misc-difference 169 /note= "encoded by GKT"
XX
WC200052135-A2.
XX
OB-SEP-2000.
XX
OI-MAR-2000; 2000WO-US005313.
XX
O2-MAR-1999; 99US-0122582P.
PR OB-DEC-1999; 99US-0169624P.
XX
{HUMA-} HUMAN GENOME SCI INC.
PA {UYJO } UNIV JOHNS HOPKINS.
PA {UYWY-} UNIV WYOMING.
XX
PI Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;
XX WPI; 2000-572178/53.
DR N-PSDB; AAM50568.
XX
PT Recombinant production of sialylated glycoproteins using cells in which
PT the expression of enzymes, e.g. sialic acid synthetase, involved in the
XX PT sialylation reaction has been altered.
PS Claim 20; Page 102-103; 144pp; English.

XX The present sequence is that of human CMP-sialic acid synthetase. The
 CC invention provides methods and recombinantly engineered cells for
 CC producing glycoproteins having sialylated oligosaccharides. The methods
 CC involve altering the expression of enzymes involved in carbohydrate
 CC processing. A claimed cell producing sialylated glycoprotein at above
 CC endogenous levels expresses at least 1 (preferably human) enzyme selected
 CC from GlcNAc-2 epimerase, an enzyme catalyzing the conversion of UDP-
 CC GlcNAc to ManNAc, sialic acid synthetase, aldolase, CMP-sialic acid
 CC synthetase and CMP-sialic acid transporter at above endogenous levels.
 CC Endogenous N-acetylglucosaminidase activity may be suppressed. A claimed
 CC method for manipulating glycoprotein in an insect cell comprises
 CC enhancing the expression of 1 of the above enzymes, and a claimed method
 CC for producing sialylated glycoproteins involves expressing a heterologous
 CC protein (especially plasminogen, transferrin, Na,K-ATPase or
 CC thyrotropin) in the insect cell, yeast, insect, fungal, plant and
 CC bacterial host cells can be engineered to produce new forms of sialylated
 CC glycoproteins, higher concentrations of sialylated glycoproteins and/or
 CC elevated concentrations of donor substrates (e.g. nucleotide sugars)
 CC required for sialylation
 XX
 SQ Sequence 434 AA;

Query Match 99.6%; Score 2258; DB 3; Length 434;
 Best Local Similarity 100.0%; Pred. No. 2.2e-223;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSVEKGATSVSNRGRPSRGRPKLQNSRGGQGRGVERPPHAAALILARGGSGKIEL 60
 Db 1 MDSVEKGATSVSNRGRPSRGRPKLQNSRGGQGRGVERPPHAAALILARGGSGKIEL 60
 QY 61 KNIKHLAGVPLIGVLRALDSGAFQSVWSTDHDEIENKAFQGVHRSSEVSKDSS 120
 Db 61 KNIKHLAGVPLIGVLRALDSGAFQSVWSTDHDEIENKAFQGVHRSSEVSKDSS 120
 QY 121 TSLDAITIEFLNXXEXDIVGNIQATXCLHPDTLQKVAEMIREEGYDSXFSVVRHQRW 180
 Db 121 TSLDAITIEFLNXXEXDIVGNIQATXCLHPDTLQKVAEMIREEGYDSXFSVVRHQRW 180
 QY 181 SEIQGVREVTPELNINPAKPRRODWDGELYENGSPFAKHLIEMGYLQGGKWHITTC 240
 Db 181 SEIQGVREVTPELNINPAKPRRODWDGELYENGSPFAKHLIEMGYLQGGKWHITTC 240
 QY 241 ELESVDIDVDIWEPIAEORVLRVYFGKEKLEIKLVNCDGLTNGHIVVSGDQKEI 300
 Db 241 ELESVDIDVDIWEPIAEORVLRVYFGKEKLEIKLVNCDGLTNGHIVVSGDQKEI 300
 QY 301 ISYDVKDAIGISLLKKSIEVRLISERACSKQTLSLKLDCMVEVSDKLAVDWEKKE 360
 Db 301 ISYDVKDAIGISLLKKSIEVRLISERACSKQTLSLKLDCMVEVSDKLAVDWEKKE 360
 QY 361 YGLCWKEVAYLIGNEVSDKLVKSLGAPADACSYAQKAVGYICKNGGRGATREFAEH 420
 Db 361 YGLCWKEVAYLIGNEVSDKLVKSLGAPADACSYAQKAVGYICKNGGRGATREFAEH 420
 QY 421 ICLLMKVNNSCKQ 434
 Db 421 ICLLMKVNNSCKQ 434

RESULT 2
 ID AAY90351 standard; protein; 434 AA.
 AC AAY90351;
 XX
 DT 04-DEC-2000 (first entry)

Human glycosylation enzyme clone HwLM34 protein sequence.
 Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
 immunotherapy; cosmetic surgery; metabolism; immune system disorder;
 haematopoietic cell deficiency; blood coagulation disorder; asthma;

KW afibrinogenaemia; blood platelet disorder; thrombocytopaenia; neoplasia;
 KW autoimmune disorder; Addison's disease; multiple sclerosis; purpura;
 KW allergic encephalomyelitis; allergic reaction; organ rejection;
 KW graft-versus-host disease; inflammation; hyperproliferative disorder;
 KW sarcoidosis; infection; gene therapy; CMP sialic acid synthetase.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 133 /note= "encoded by YAT"
 FT Misc-difference 136 /note= "encoded by GKT"
 FT Misc-difference 147 /note= "encoded by YCA"
 FT Misc-difference 169 /note= "encoded by GKT"
 PN WO200052136-A2.
 XX
 PD 08-SEP-2000.
 XX
 PF 01-MAR-2000; 2000WO-US005325.
 PR 02-MAR-1999; 99US-0122409P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Colatan TA;
 XX
 DR WPI; 2000-572179/53.
 DR N-PSDB; AAA37762.
 XX
 DR New human glycosylation enzymes cytidine 5'-monophosphate sialic acid
 FT synthetase, sialic acid synthetase and aldolase and nucleic acids
 PT encoding the proteins for treating e.g., immune system disorders,
 PT microbial diseases.
 XX
 PS Claim 12; Page 106-108; 1159p; English.
 XX
 CC This sequence represents a human glycosylation enzyme clone of the
 CC invention, designated HwLM34. This protein clone is a CMP sialic acid
 CC synthetase. The sequences are useful as reagents for the differential
 CC identification of the tissues or cell types present in a biological
 CC sample, as immunological probes, for treating a disease or condition
 CC and/or treatment of disorders involving aberrant glycolysis, e.g. cramps,
 CC myoglobinuria, and as tumour marker and/or immunotherapy targets. They
 CC may also be used to differentiate, proliferate and attract cells leading
 CC to the regeneration of tissues, to modulate mammalian characteristics
 CC (e.g. in cosmetic surgery) or mammalian metabolism affecting catabolism,
 CC anabolism processing, utilisation and energy storage, to change a
 CC mammal's mental state by influencing and as a food additive or
 CC preservative. The proteins can be used to assay protein levels in a
 CC sample, as a marker or detector of an immune system disorder, to inhibit
 CC cytokine activity, and as a vaccine. They may further be used to treat
 CC immune system or of haematopoietic cell deficiencies or disorders, blood
 CC coagulation disorders (e.g. afibrinogenaemia), blood platelet disorders
 CC (e.g. thrombocytopaenia), wounds resulting from trauma or surgery,
 CC autoimmune disorders (e.g. Addison's disease, multiple sclerosis,
 CC allergic encephalomyelitis), allergic reactions (e.g. asthma, organ
 CC rejection, graft-versus-host disease, inflammation, hyperproliferative
 CC disorders (e.g. neoplasia, purpura, sarcoidosis), diseases caused by
 CC viruses (e.g. hepatitis, meningitis, AIDS), bacteria and fungi (include
 CC e.g. tuberculosis, conjunctivitis, sepsis, typhoid, chlamydia,
 CC cellulitis), and diseases caused by parasites (e.g. amoebiasis,
 CC coccidiosis, leishmaniasis, scabies, malaria, toxoplasmosis)
 XX
 SQ Sequence 434 AA;

Query Match 99.6%; Score 2258; DB 3; Length 434;
 Best Local Similarity 100.0%; Pred. No. 2.2e-223;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVNPRGRPSRGRPPKLRNSRGQGRGVEKPPHLLAALILARGSGKIPL 60
 DB 1 MDSVEKGAATSVNPRGRPSRGRPPKLRNSRGQGRGVEKPPHLLAALILARGSGKIPL 60
 QY 61 KNKHLAGVPLIGWVLRALDGAFOQVWVSTDHDEIENVAQFGAQQVHRRSEVSKDSS 120
 DB 61 KNKHLAGVPLIGWVLRALDGAFOQVWVSTDHDEIENVAQFGAQQVHRRSEVSKDSS 120
 QY 121 TSLSDAIEFLNYXNEXDIVGNIGATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHOFRW 180
 DB 121 TSLSDAIEFLNYXNEXDIVGNIGATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHOFRW 180
 QY 181 SEIQKGVREVTPEPLNPNPAKRPRQDWDGELYENGSGFYFAKRLHIEMGYLQGGKWHHTKC 240
 DB 181 SEIQKGVREVTPEPLNPNPAKRPRQDWDGELYENGSGFYFAKRLHIEMGYLQGGKWHHTKC 240
 QY 241 ELEHSVDIDVDIDWPIAEOQVLYGYFGKELKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
 DB 241 ELEHSVDIDVDIDWPIAEOQVLYGYFGKELKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
 QY 301 ISYDVKDAIGISLLKSGIEVRLISERACKOTLSLKLDCCKEVSVDKLVAVDDEWKE 360
 DB 301 ISYDVKDAIGISLLKSGIEVRLISERACKOTLSLKLDCCKEVSVDKLVAVDDEWKE 360
 QY 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQAQVYICKNGGGRGAIRFAEH 420
 DB 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQAQVYICKNGGGRGAIRFAEH 420
 QY 421 ICLMEKVNNSCQK 434
 DB 421 ICLMEKVNNSCQK 434

RESULT 3

AAB84682
 ID AAB84682 standard; protein; 434 AA.
 AC AAB84682;
 XX
 DT 17-SEP-2001 (first entry)
 DE Nucleotide sequence of a cytidine monophosphate-sialic acid synthetase.
 XX
 KW Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;
 KW cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;
 KW sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;
 KW vaccine.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 133 /note= "unspecified residue encoded by YAT"
 FT Misc-difference 136 /note= "unspecified residue encoded by GKT"
 FT Misc-difference 147 /note= "unspecified residue encoded by YCA"
 FT Misc-difference 169 /note= "unspecified residue encoded by GKT"
 FT
 XX
 PN W020312492-A1.
 XX
 PD 14-JUN-2001.
 XX
 PF 07-DEC-2003; 2000W0-U5033136.
 XX
 PF 09-DEC-1993; 99US-0169839P.
 PR
 XX {HUMA-} HUMAN GENOME SCI INC.
 PA {UYJO } UNIV JOHNS HOPKINS.
 PA {UTEM } UNIV TEMPLE.
 PA {UYWY-} UNIV WYOMING.

XX
 PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Falter K;
 XX Jarvis D;
 XX WPZ; 2001-441575/47.
 DR N-PSDB; AAH28457.
 XX
 FT Cells producing cytidine monophosphate-sialic acid and sialylated
 PT glycoprotein above endogenous levels for production of vaccines and
 PT therapeutics.
 XX
 PS Claim 20; Fig 30; 182pp; English.
 XX
 CC The specification describes a method for manipulating carbohydrate
 CC processing pathways in cells of interest. The methods are used to
 CC manipulate multiple pathways involved with the sialylation reaction by
 CC using recombinant DNA technology and substrate feeding approaches to
 CC enable the production of sialylated glycoproteins in the cells. The
 CC sialylation process involves the post-translational addition of the donor
 CC substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific
 CC acceptor carbohydrate. The cells express at least one enzyme, selected
 CC from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,
 CC CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The
 CC cells are useful for producing complex sialylated glycoproteins in cells
 CC of interest, especially insect cells. Glycoproteins containing sialylated
 CC oligosaccharides are useful as vaccines, therapeutics and diagnostic
 CC tools. Cells producing complex sialylated glycoproteins are useful for
 CC enhancing the value of heterologous expression systems and increasing the
 CC application of heterologous cell expression products as vaccines,
 CC therapeutics and diagnostic tools as well as increasing the variety of
 CC heterologous proteins that can be produced and lowering biotechnology
 CC production costs. The present sequence represents a human CMP-SA
 CC synthetase, which is used in the method of the invention
 XX
 SQ Sequence 434 AA;

Query Match 99.5%; Score 2258; DB 4; Length 434;
 Best Local Similarity 100.0%; Pred. No. 2.2e-223;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDSVEKGAATSVNPRGRPSRGRPPKLRNSRGQGRGVEKPPHLLAALILARGSGKIPL 60
 DB 1 MDSVEKGAATSVNPRGRPSRGRPPKLRNSRGQGRGVEKPPHLLAALILARGSGKIPL 60
 QY 61 KNKHLAGVPLIGWVLRALDGAFOQVWVSTDHDEIENVAQFGAQQVHRRSEVSKDSS 120
 DB 61 KNKHLAGVPLIGWVLRALDGAFOQVWVSTDHDEIENVAQFGAQQVHRRSEVSKDSS 120
 QY 121 TSLSDAIEFLNYXNEXDIVGNIGATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHOFRW 180
 DB 121 TSLSDAIEFLNYXNEXDIVGNIGATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHOFRW 180
 QY 181 SEIQKGVREVTPEPLNPNPAKRPRQDWDGELYENGSGFYFAKRLHIEMGYLQGGKWHHTKC 240
 DB 181 SEIQKGVREVTPEPLNPNPAKRPRQDWDGELYENGSGFYFAKRLHIEMGYLQGGKWHHTKC 240
 QY 241 ELEHSVDIDVDIDWPIAEOQVLYGYFGKELKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
 DB 241 ELEHSVDIDVDIDWPIAEOQVLYGYFGKELKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
 QY 301 ISYDVKDAIGISLLKSGIEVRLISERACKOTLSLKLDCCKEVSVDKLVAVDDEWKE 360
 DB 301 ISYDVKDAIGISLLKSGIEVRLISERACKOTLSLKLDCCKEVSVDKLVAVDDEWKE 360
 QY 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQAQVYICKNGGGRGAIRFAEH 420
 DB 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQAQVYICKNGGGRGAIRFAEH 420
 QY 421 ICLMEKVNNSCQK 434
 DB 421 ICLMEKVNNSCQK 434

RESULT 4
 AA026545
 ID AA026545 standard; protein; 434 AA.
 XX
 AC AA026545;
 DT 06-MAR-2003 (first entry)
 XX
 DE Human CMP-SA synthetase protein.
 XX
 KW Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;
 KW GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)NAC; sialic acid;
 KW synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;
 KW transporter; sialylated glycoprotein; human.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 133 /label= His, Tyr
 FT /note= 'Amino acid residue is encoded by YAT'
 FT Misc-difference 136 /label= Gly, Val
 FT /note= 'Amino acid residue is encoded by GKT'
 FT Misc-difference 147 /label= Pro, Ser
 FT /note= 'Amino acid residue is encoded by YCA'
 FT Misc-difference 169 /label= Gly, Val
 FT /note= 'Amino acid residue is encoded by GKT'
 XX
 US2002142386-A1.
 XX
 03-OCT-2002.
 XX
 16-AUG-2002; 2001US-00930440.
 XX
 02-MAR-1999; 99US-0122582P.
 PR 08-DEC-1999; 99US-0169624P.
 PR 25-AUG-2000; 2000US-0227579P.
 XX
 (BETE/) BETENBAUGH M J.
 PA (LAWR/) LAWRENCE S.
 PA (LEEY/) LEE Y C.
 PA (COLE/) COLEMAN T A.
 XX
 Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;
 XX
 W2; 2003-102519/09.
 DR N-PSDB; AAL53992.
 XX
 Manipulating glycoprotein production in insect cell, involves enhancing
 expression of enzymes involved in carbohydrate processing pathway such as
 N-acetylglucosamine-2 epimerase or sialic acid synthetase.
 XX
 Claim 20; Fig 30; 88pp; English.
 XX
 The invention relates to a novel method for manipulating glycoprotein
 production in an insect cell comprising enhancing expression of an
 enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
 catalysing conversion of UDP-GlcNAc to mannose (Man)NAC, sialic acid
 synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)
 synthetase or CMP-SA transporter, where the expression of each enzyme is
 enhanced to above endogenous levels. The novel method is useful for
 manipulating glycoprotein production in an insect cell. Further methods
 of the invention are useful for producing sialylated glycoprotein. The
 sialylated glycoprotein produced by the above mentioned methods are
 useful as pharmaceutical compositions, vaccines, diagnostics and
 therapeutics. This sequence represents the human CMP-SA synthetase
 protein of the invention
 XX
 Sequence 434 AA;

Query Match 99.6%; Score 2258; DB 6; Length 434;
 Best Local Similarity 100.0%; Pred. No. 2.2e-223; Indels 0; Gaps 0;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDSVEKGAATSVNPRGRSPRPPKLRNRRGGGGRGVEKPPHDAALILARGSGKIGPL 60
 XX
 DB 1 MDSVEKGAATSVNPRGRSPRPPKLRNRRGGGGRGVEKPPHDAALILARGSGKIGPL 60
 XX
 QY 61 KNIKHLAGVPLIGVLRALDSGAFQSVVSTDDHDEIENVAQFCAQVHRSSSEYKSS 120
 XX
 DB 61 KNIKHLAGVPLIGVLRALDSGAFQSVVSTDDHDEIENVAQFCAQVHRSSSEYKSS 120
 XX
 QY 121 TSLDAITIEFLNYXNEXDIVGNIQATSKLHPTDLQKVAEMIREEGYDSKFSVVRHQFRW 180
 XX
 DB 121 TSLDAITIEFLNYXNEXDIVGNIQATSKLHPTDLQKVAEMIREEGYDSKFSVVRHQFRW 180
 XX
 QY 181 SEIQGVREVTPEPLNPAKPRRQDWDGELYENGSPYFAKHLEIEMGYLQGGKWHITKC 240
 XX
 DB 181 SEIQGVREVTPEPLNPAKPRRQDWDGELYENGSPYFAKHLEIEMGYLQGGKWHITKC 240
 XX
 QY 241 ELEHSVDIDVDIWDPIAEQVRVRYGFKSEKELKELILVCMIDGCLTNGHIYVSGDQKEI 300
 XX
 DB 241 ELEHSVDIDVDIWDPIAEQVRVRYGFKSEKELKELILVCMIDGCLTNGHIYVSGDQKEI 300
 XX
 QY 301 ISYDVKDAIGISILKKGIEVRLISERACSKQTSSLKLDCKMEVSVSKLAVDVEWKE 360
 XX
 DB 301 ISYDVKDAIGISILKKGIEVRLISERACSKQTSSLKLDCKMEVSVSKLAVDVEWKE 360
 XX
 QY 361 MGLCWKEVAYLGNESDDECLKRVGLSGHPADACSYAQKAVGYICKNGRGRAIFREFAH 420
 XX
 DB 361 MGLCWKEVAYLGNESDDECLKRVGLSGHPADACSYAQKAVGYICKNGRGRAIFREFAH 420
 XX
 QY 421 ICLMEKVNNSCQK 434
 XX
 DB 421 ICLMEKVNNSCQK 434
 XX
 RESULT 5
 AAB28673
 ID AAB28673 standard; protein; 434 AA.
 XX
 AC AAB28673;
 XX
 DT 13-FEB-2001 (first entry)
 XX
 DE Human carbohydrate-modifying enzyme Incyte ID No: 000422CD1.
 XX
 KW Human; carbohydrate-modifying enzyme; CME; antidiabetic;
 KW immunosuppressive; anti-HIV; antiinflammatory; antianaemic;
 KW antiasthmatic; antiarteriosclerotic; antithyroid; hepatotropic;
 KW nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic;
 KW antiarthritic; antiposoriatic; uropathic; ophthalmological;
 KW dermatological; antiulcer; cytosstatic; virucide; antibacterial;
 KW fungicide; protozoacide; tranquilliser; vulnerary; diabetes;
 KW autoimmune disorder; inflammatory disorder; infection.
 XX
 OS Homo sapiens.
 XX
 PK W0200063351-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 20-APR-2000; 2000WO-US010882.
 XX
 PR 21-APR-1999; 99US-0130383P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lal P, Yue H, Tang Yt, Hillman JL, Baughn MR, Yang J;
 XX
 DR WPI; 2000-672723/65.
 DR N-PSDB; AAC65392.
 XX

PT Novel carbohydrate modifying enzyme polypeptides and polynucleotides for
PT diagnosis, treatment, and prevention of carbohydrate metabolism
XX disorders, autoimmune/inflammatory disorders, and cancer.
XX
XX Claim 1; Page 66-67; 75pp; English.
XX
XX The present sequence is a human carbohydrate-modifying enzyme (CME). CME
XX polynucleotides and polypeptides are useful for treating and diagnosing
XX diseases associated with CME such as diabetes, autoimmune/inflammatory
XX disorders such as AIDS, Addison's disease, adult respiratory distress
XX syndrome, allergies, anaemia, asthma, atherosclerosis, autoimmune
XX thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's
XX disease, emphysema, erythroblastosis fetalis, glomerulonephritis, Good
XX pasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis,
XX multiple sclerosis, myasthenia gravis, osteoarthritis, osteoporosis,
XX pancreatitis, polymyositis, psoriasis, Reiter's syndrome, arthritis,
XX scleroderma, Sjogren's syndrome, systemic lupus erythematosus, ulcerative
XX colitis, uveitis, Werner syndrome, complications of cancer,
XX haemodialysis, and extracorporeal circulation, viral, bacterial, fungal
XX parasitic, protozoal, and helminthic infections, trauma, or cancer. CME,
XX or its catalytic or immunogenic fragment, is useful for drug screening
XX
XX Sequence 434 AA;
XX
XX Query Match 96.6%; Score 2188; DB 3; Length 434;
XX Best Local Similarity 97.0%; Pred. No. 3.7e-216;
XX Matches 421; Conservative 1; Mismatches 12; Indels 0; Gaps 0;
XX
XX Qy 1 MDSVEKGAATSVSNPRGRPPKLRNRRGGQGRGVEKPPHAAALILARGSGSGIPL 60
XX Db 1 MDSVEKGAATSVSNPRGRPPKLRNRRGGQGRGVEKPPHAAALILARGSGSGIPL 60
XX
XX Qy 61 KNIXHLAGVPLIGWVLAALDSGAFQSVVSTHDEIENVAKQGAQVHRSSESVKSS 120
XX Db 61 KNIXHLAGVPLIGWVLAALDSGAFQSVVSTHDEIENVAKQGAQVHRSSESVKSS 120
XX
XX Qy 121 TSLDAITFEFLNYXNEXDIVGNIQATSCXCLHPTDLOKVAEMIREGSDYDFSVVRHQRW 180
XX Db 121 TSLDAITFEFLNYHNEVDIVGNIQATSPCLHPTDLOKVAEMIREGSDYDFSVVRHQRW 180
XX
XX Qy 181 SEIQGVRETEPLNPAKPRQDMDGELYENGFSYFAKRLIEMGYLQGGKWHHTKC 240
XX Db 181 SEIQGVRETEPLNPAKPRQDMDGELYENGFSYFAKRLIEMGYLQGGKWHHTKC 240
XX
XX Qy 241 ELESVDIDVDIDWPIAEORVIRYGFGEKELKILVCMIDCLTNHGYVSGDQKEI 300
XX Db 241 RAEHSDVDIDVDIDWPIAEORVIRYGFGEKELKILVCMIDCLTNHGYVSGDQKEI 300
XX
XX Qy 301 ISYDVDAIGISLLKSGIEVRLISERACSKQTISSLKLDCKMEVSDKLAVDWEKKE 360
XX Db 301 ISYDVDAIGISLLKSGIEVRLISERACSKQTISSLKLDCKMEVSDKLAVDWEKKE 360
XX
XX Qy 361 MGLCKEVAVLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGSGALREFAEH 420
XX Db 361 MGLCKEVAVLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGSGALREFAEH 420
XX
XX Qy 421 ICLLMEKWNNSCQK 434
XX Db 421 ICLLMEKWNNSCQK 434
XX
XX RESULT 6
XX ID AM39319 standard; protein; 434 AA.
XX AC AM39319;
XX XX
XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 2464.
XX EN Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
XX
XX WC200153312-AL.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
XX 21-JAN-2000; 2000US-00488725.
XX 25-APR-2000; 2000US-00522317.
XX 20-JUN-2000; 2000US-00598042.
XX 19-JUL-2000; 2000US-00620312.
XX 03-AUG-2000; 2000US-00653450.
XX 14-SEP-2000; 2000US-00662191.
XX 19-OCT-2000; 2000US-00693036.
XX 29-NOV-2000; 2000US-00727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX N-PSDB; AAL58475.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2464; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
XX encoded polypeptides (AA38642-AA42213) with neurotropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification
XX
XX Sequence 434 AA;
XX
XX Query Match 96.6%; Score 2188; DB 4; Length 434;
XX Best Local Similarity 97.0%; Pred. No. 3.7e-216;
XX Matches 421; Conservative 1; Mismatches 12; Indels 0; Gaps 0;
XX
XX Qy 1 MDSVEKGAATSVSNPRGRPPKLRNRRGGQGRGVEKPPHAAALILARGSGSGIPL 60
XX Db 1 MDSVEKGAATSVSNPRGRPPKLRNRRGGQGRGVEKPPHAAALILARGSGSGIPL 60
XX
XX Qy 61 KNIXHLAGVPLIGWVLAALDSGAFQSVVSTHDEIENVAKQGAQVHRSSESVKSS 120
XX Db 61 KNIXHLAGVPLIGWVLAALDSGAFQSVVSTHDEIENVAKQGAQVHRSSESVKSS 120
XX
XX Qy 121 TSLDAITFEFLNYXNEXDIVGNIQATSCXCLHPTDLOKVAEMIREGSDYDFSVVRHQRW 180
XX Db 121 TSLDAITFEFLNYHNEVDIVGNIQATSPCLHPTDLOKVAEMIREGSDYDFSVVRHQRW 180
XX
XX Qy 181 SEIQGVRETEPLNPAKPRQDMDGELYENGFSYFAKRLIEMGYLQGGKWHHTKC 240

Db 181 SEIQGVREVTPELNINPAKRRPRQDWDGEVYENGSGFYFAKPHLIEMGVLOGGKMAVYEM 240

Qy 241 ELEHSVDIDVDIDWDPIAQRVLYRYFGYFGKEKKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300

Db 241 RAHESVDIDVDIDWDPIAQRVLYRYFGYFGKEKKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300

Qy 301 ISYDVKDAIGISLLKKSGLIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDWEWKE 360

Db 301 ISYDVKDAIGISLLKKSGLIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDWEWKE 360

Qy 361 MGLCWKEVAYLNGNEVSDDECLKRVLSGAPADACSYAQKAVGYICKNGRGGAIRFAEH 420

Db 361 MGLCWKEVAYLNGNEVSDDECLKRVLSGAPADACSYAQKAVGYICKNGRGGAIRFAEH 420

Qy 421 ICLLMKVNNSCQK 434

Db 421 ICLLMKVNNSCQK 434

RESULT 7

AAB95213

ID AAB95213 standard; protein; 434 AA.

XX

AC AAB95213;

DT 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:17325.

XX

KW Hmat; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

CS Homo sapiens.

XX

PN EPI074617-A2.

XX

PS 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-00116126.

XX

PR 29-JUL-1999; 98JP-00248016.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI; 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX

PS Claim 8; SEQ ID NO 17325; 2537pp + Sequence Listing; English.

XX

CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs easily without any specialised methods. AAB03166 to AAB13628 and AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to AAB95993 represent human amino acid sequences; and AAB13629 to AAB13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

CC particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAB03166 to AAB13628 and AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to AAB95993 represent human amino acid sequences; and AAB13629 to AAB13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX

SQ Sequence 434 AA;

Query Match 96.6%; Score 2188; DB 4; Length 434;

Best Local Similarity 97.0%; Pred. No. 3.7e-216;

Matches 421; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MDSVEKGAATSVSNPRGRPPKLRNSRGGGGRGVEKPPHLLAALILARGSGKGIPL 60

Db 1 MDSVEKGAATSVSNPRGRPPKLRNSRGGGGRGVEKPPHLLAALILARGSGKGIPL 60

Qy 61 KNIKHLAGVPLIGVLRALDSGAFQSVVSTDHDEIENVAQFQAVHRRSEVSKDSS 120

Db 61 KNIKHLAGVPLIGVLRALDSGAFQSVVSTDHDEIENVAQFQAVHRRSEVSKDSS 120

Qy 121 TSLEDAIIEFLNYXNEXDIVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180

Db 121 TSLEDAIIEFLNYHNEVDIVGNIQATSPCLHPTDLOKVAEMIREEGYDSVFSVVRHQFRW 180

Qy 181 SEIQGVREVTPELNINPAKRRPRQDWDGEVYENGSGFYFAKPHLIEMGVLOGGKMAVYEM 240

Db 181 SEIQGVREVTPELNINPAKRRPRQDWDGEVYENGSGFYFAKPHLIEMGVLOGGKMAVYEM 240

Qy 241 ELEHSVDIDVDIDWDPIAQRVLYRYFGYFGKEKKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300

Db 241 RAHESVDIDVDIDWDPIAQRVLYRYFGYFGKEKKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300

Qy 301 ISYDVKDAIGISLLKKSGLIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDWEWKE 360

Db 301 ISYDVKDAIGISLLKKSGLIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDWEWKE 360

Qy 361 MGLCWKEVAYLNGNEVSDDECLKRVLSGAPADACSYAQKAVGYICKNGRGGAIRFAEH 420

Db 361 MGLCWKEVAYLNGNEVSDDECLKRVLSGAPADACSYAQKAVGYICKNGRGGAIRFAEH 420

Qy 421 ICLLMKVNNSCQK 434

Db 421 ICLLMKVNNSCQK 434

RESULT 8

AAB43931

ID AAB43931 standard; protein; 448 AA.

XX

AC AAB43931;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human cancer associated protein sequence SEQ ID NO:1376.

XX

KW Human; cancer associated gene; cancer antigen; detection; cancer; diagnosis; cytostatic; proliferative; vulnerable; immunomodulator; antidiabetic; antiasumatic; antirheumatic; antithrombotic; antiviral; antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac; dermatological; neuroprotective; thrombolytic; coagulant; nocotropic; vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation; immune disorder; haematopoietic cell disorder; autoimmune disorder; allergic reaction; graft versus host disease; organ rejection; haemostatic; thrombolytic; cardiovascular disorder; infection; neurological disease; drug screening.

XX

OS Homo sapiens.

XX

PN WO20005530-A1.

XX

PD 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US005882.
XX 12-MAR-1999; 99US-0124270P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rose: CA, Ruben SM;
XX WPI: 2000-587533/55.
XX N-PSDB; AAC78140.
XX Novel isolated nucleic acids comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer.
XX Claim 11; Page 2041-2042; 2352pp; English.
XX AAC77607 to AAC78448 encode the human cancer associated proteins given in
XX AAB43398 to AAB44239. The proteins can have activities based on the
XX tissues and cells the genes are expressed in. Example of activities
XX include: cytostatic; proliferative; vulnarary; immunomodulator;
XX antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
XX antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
XX dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
XX nootropic; vasotropic; antipsoriatic and antiangiogenic. The
XX polynucleotides and polypeptides can be used for preventing, treating or
XX ameliorating medical conditions and diagnosing pathological conditions.
XX Polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX the present invention may be used to treat immune disorders by activating
XX or inhibiting the proliferation, differentiation or mobilisation of
XX immune cells, to treat disorders of haematopoietic cells, autoimmune
XX disorders, allergic reactions, graft versus host disease and organ
XX rejection, modulate haemostatic or thrombolytic activity, modulate
XX inflammation, cancers, cardiovascular disorders, neurological disease and
XX bacterial or viral infections. The peptides, nucleotides, antibodies,
XX agonists and antagonists may be also be used in drug screens. AAC78449 to
XX AAC78457 and AAB44240 represent sequences used in the exemplification of
XX the present invention
XX Sequence 448 AA;
XX
XX Query Match 96.6%; Score 2188; DB 3; Length 448;
XX Best Local Similarity 97.0%; Pred. No. 3.9e-216;
XX Matches 421; Conservative 1; Mismatches 12; Indels 0; Gaps 0;
XX
XX 1 MDSVEKGAATSVNPRGRPRGRPPKLRNSRGQGRGVEKPEHLAAILARGGSKIPL 60
XX 15 MDSVEKGAATSVNPRGRPRGRPPKLRNSRGQGRGVEKPEHLAAILARGGSKIPL 74
XX
XX 61 KNIKHLAGVPLIGWTLRAALDSGAFQSVVSTVDHDEIENVAKQFGAQRHRSSEVSKDS 120
XX 75 KNIKHLAGVPLIGWTLRAALDSGAFQSVVSTVDHDEIENVAKQFGAQRHRSSEVSKDS 134
XX
XX 121 TSDALIEFLNYNEVDIVNIGICATSCLEPTDLQKVAEMIREGVDYXSVVRHOFRW 180
XX 135 TSDALIEFLNYNEVDIVNIGICATSCLEPTDLQKVAEMIREGVDYXSVVRHOFRW 194
XX
XX 181 SETQGVRETEPLNPAKRRQDWDGELYENGSEYFAKRLHIEGVLQGGKWHITTC 240
XX 195 SE-QGVRETEPLNPAKRRQDWDGELYENGSEYFAKRLHIEGVLQGGKWHITTC 254
XX
XX 241 ELBHSVDIDVDIWPITAEQRLVRYGFGKGLKEIKLLVNDGCLTNHGH-YVSGDQKEI 300
XX 255 RAESHVDIDVDIWPITAEQRLVRYGFGKGLKEIKLLVNDGCLTNHGH-YVSGDQKEI 314
XX
XX 301 ISYDVNDKDAIGSLIKKSGIEVRLISERACSKQTLSLKLCKEVEVSVDKLAVDENRKE 360
XX 315 ISYDVNDKDAIGSLIKKSGIEVRLISERACSKQTLSLKLCKEVEVSVDKLAVDENRKE 374
XX
XX 361 MGLCWKEVAYLGNVEVDEELCKEYVLSGAPADACSYAKQAVGYICKNGGGAIRPFAEH 420
XX 375 MGLCWKEVAYLGNVEVDEELCKEYVLSGAPADACSYAKQAVGYICKNGGGAIRPFAEH 434

QY 421 ICLLMKVNNSCOK 434
DB 435 ICLLMKVNNSCOK 448

RESULT 3

AAB43155
ID AAB43165 standard; protein; 434 AA.
XX
AC AAB43165;
XX
XX
DT 08-FEB-2001 (first entry)
XX
DE Hunan ORFX ORF2929 polypeptide sequence SEQ ID NO:5858.
XX

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnarary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antinflammatory;
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX actinaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive.
XX

OS Homo sapiens.
XX

PN WO200058473-A2.
XX

PD 05-OCT-2000.
XX

XX 31-MAR-2000; 2000WO-US008521.
XX

PR 31-MAR-1999; 99US-0127607P.
XX

PR 02-APR-1999; 99US-0127636P.
XX

PR 05-APR-1999; 99US-0127728P.
XX

PR 30-MAR-2000; 2000US-00540763.
XX

XX (CURA-) CURAGEN CORP.
XX

XX Shimkets RA, Leach M;
XX

XX WPI: 2000-602362/57.
XX

XX N-PSDB; AAC77374.
XX

XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX

XX Claim 11; Page 5028-5029; 5507pp; English.
XX

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnarary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
XX anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antinflammatory; antibacterial;
XX antiviral; antifungal; antirheumatic; antithyroid; and actinaemic. The
XX sequences can be used for determining the presence of or predisposition
XX to, or preventing or treating pathological conditions associated with an
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancers, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency

CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, anti-inflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 SQ Sequence 434 AA;
 Query Match 96.2%; Score 21491; DB 3; Length 434;
 Best Local Similarity 96.8%; Pred. No. 26-215;
 Matches 420; Conservative 1; Mismatches 13; Indels 0; Gaps 0;
 QY 1 MDSVEKGAATSVSNPRGPRPPKLRNRRGGQGRGVEKPPHLLAAILARGSGKGIP 60
 Db 1 MDSVEKGAATSVSNPRGPRPPKLRNRRGGQGRGVEKPPHLLAAILARGSGKGIP 60
 QY 61 KNIKHLAVPLIGWVLRRAALDSGAFQSVVSTHDEIENVAQFGAQRHRSSEVSKDS 120
 Db 61 KNIKHLAVPLIGWVLRRAALDSGAFQSVVSTHDEIENVAQFGAQRHRSSEVSKDS 120
 QY 121 TSLDAIIEFLNXXEXDIVGNICATSCGLHPTDLOKVAEMIREEGYDSKFSVVRHQRW 180
 Db 121 TSLDAIIEFLNXXEXDIVGNICATSCGLHPTDLOKVAEMIREEGYDSKFSVVRHQRW 180
 QY 181 SEIQKGVRETEPLNPAKPRRQDWDGELYENGFFAKHLLIEMGYLOGGKWHITKC 240
 Db 181 SEIQKGVRETEPLNPAKPRRQDWDGELYENGFFAKHLLIEMGYLOGGKWHITKC 240
 QY 241 ELEHSDVIDDIDWPIAQORVLRVYFGKEKKEIKLLVCNIDGCLTNHIVSGDQKEI 300
 Db 241 ELEHSDVIDDIDWPIAQORVLRVYFGKEKKEIKLLVCNIDGCLTNHIVSGDQKEI 300
 QY 301 ISYDVKDAIGSLKKSGLIEVRLISERACSKOTLSSKLDCKMEVSDKLAVDWBRKE 360
 Db 301 ISYDVKDAIGSLKKSGLIEVRLISERACSKOTLSSKLDCKMEVSDKLAVDWBRKE 360
 QY 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQKAVGYICKNGGRGAIRFAEH 420
 Db 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQKAVGYICKNGGRGAIRFAEH 420
 QY 421 ICLLMEKVNNSQK 434
 Db 421 ICLLMEKVNNSQK 434
 RESULT 10
 AAM41105
 ID AAM41105 standard; protein; 461 AA.
 AC AAM41105;
 XX AAM41105;
 DT 22-OCT-2001 (first entry)
 XX Human polypeptide SEQ ID NO 6036.
 DE Human polypeptide SEQ ID NO 6036.
 XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX Homo sapiens.
 OS Homo sapiens.
 EN WO200153312-A1.
 XX WO200153312-A1.
 FL 26-JUL-2001.
 XX 26-DEC-2000; 2000WO-US034263.
 XX 23-DEC-1999; 95US-00471275.
 XX 21-JAN-2000; 2000US-00488725.
 XX 25-APR-2000; 2000US-00552317.
 XX 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662151.
 PR 19-OCT-2000; 2000US-00693036.
 PR 25-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Targ YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AAL60261.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 PS Example 2; SEQ ID NO 6036; 10078pp; English.
 CC
 CC The invention relates to human nucleic acids (AA157798-AA161369) and the
 CC encoded polypeptides (AA38642-AA42213) with nontropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX
 SQ Sequence 461 AA;
 Query Match 94.8%; Score 2149; DB 4; Length 461;
 Best Local Similarity 97.0%; Pred. No. 4.3e-212;
 Matches 414; Conservative 1; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MDSVEKGAATSVSNPRGPRPPKLRNRRGGQGRGVEKPPHLLAAILARGSGKGIP 60
 Db 15 MDSVEKGAATSVSNPRGPRPPKLRNRRGGQGRGVEKPPHLLAAILARGSGKGIP 74
 QY 61 KNIKHLAVPLIGWVLRRAALDSGAFQSVVSTHDEIENVAQFGAQRHRSSEVSKDS 120
 Db 75 KNIKHLAVPLIGWVLRRAALDSGAFQSVVSTHDEIENVAQFGAQRHRSSEVSKDS 134
 QY 121 TSLDAIIEFLNXXEXDIVGNICATSCGLHPTDLOKVAEMIREEGYDSKFSVVRHQRW 180
 Db 135 TSLDAIIEFLNXXEXDIVGNICATSCGLHPTDLOKVAEMIREEGYDSKFSVVRHQRW 194
 QY 181 SEIQKGVRETEPLNPAKPRRQDWDGELYENGFFAKHLLIEMGYLOGGKWHITKC 240
 Db 195 SEIQKGVRETEPLNPAKPRRQDWDGELYENGFFAKHLLIEMGYLOGGKWHITKC 254
 QY 241 ELEHSDVIDDIDWPIAQORVLRVYFGKEKKEIKLLVCNIDGCLTNHIVSGDQKEI 300
 Db 255 RAHSDVIDDIDWPIAQORVLRVYFGKEKKEIKLLVCNIDGCLTNHIVSGDQKEI 314
 QY 301 ISYDVKDAIGSLKKSGLIEVRLISERACSKOTLSSKLDCKMEVSDKLAVDWBRKE 360
 Db 315 ISYDVKDAIGSLKKSGLIEVRLISERACSKOTLSSKLDCKMEVSDKLAVDWBRKE 374
 QY 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQKAVGYICKNGGRGAIRFAEH 420
 Db 375 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQKAVGYICKNGGRGAIRFAEH 434
 QY 421 ICLLMEK 427
 Db 435 ICLLMEK 441

RESULT 11
 ID ADB82754 standard; protein; 245 AA.
 XX AC ADB82754;
 XX DT 04-DEC-2003 (first entry)
 XX DE Human protein sequence useful for the treatment of cancer (SeqID 1535).
 XX KW human; prostate; cancer; cytostatic; gene therapy; vaccine;
 XX KW immune response.
 XX OS Homo sapiens.
 XX PN WO2003050236-A2.
 XX PD 19-JUN-2003.
 XX PF 04-SEP-2002; 2002WC-US028214.
 XX PR 07-DEC-2001; 2001US-00012697.
 XX PA (CHIR) CHIRON CORP.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;
 PI Crkvenjakov R, Dickson M, Drmanac S, Labat I, Leshkowitz D, Kita D;
 PI Garcia V, Jones LW, Stache-Crain B, Scott EM;
 XX W21; 2003-513972/48.
 XX New polynucleotides derived from human prostate, useful for modulating
 immune response to prevent or treat cancer.
 XX Claim 18; SEQ ID NO 1535; 188pp; English.
 XX This invention relates to novel isolated polynucleotides of human origin,
 particularly isolated from the human prostate. Specifically, it refers to
 the diagnostics and therapeutics comprising these novel human
 polynucleotides, and includes the derived probes, antisense
 oligonucleotides, and antibodies thereof. The identification of these
 human prostate genes that can inhibit tumour growth is useful for
 understanding the progression and nature of complex diseases such as
 cancer, and hence they are important in the drug discovery process. The
 present invention describes these polynucleotides and encoded
 polypeptides as exhibiting cytostatic activity, and through gene therapy
 and/or vaccines they can be used to modulate the immune response for the
 prevention or treatment of cancers, particularly of the prostate, but
 also for breast, lung and colon cancer. This polypeptide sequence is a
 human protein sequence useful for the treatment of cancer, used in an
 exemplification of the invention. NOTE: These sequences are not given in
 the specification but are provided on the WIFO website.
 XX SQ Sequence 245 AA;
 Query Match 51.8%; Score 1173; DB 7; Length 245;
 Best Local Similarity 95.7%; Pred. No. 6.8e-112;
 Matches 225; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
 QY 137 DIVGNTQATSKLHPDLOKVAEMIREEGYDSFVSFVRRHQFWSHIOKGVREVTPLNL 196
 DB 3 DIVGNTQATSKLHPDLOKVAEMIREEGYDSFVSFVRRHQFWSHIOKGVREVTPLNL 62
 QY 197 NPAKRPRRQDMDGELYENGSGFYFAKHLHMGVLYOGGKWHITTKCELEHSDVDIDWDWI 256
 DB 63 NPAKRPRRQDMDGELYENGSGFYFAKHLHMGVLYOGGKWHITTKCELEHSDVDIDWDWI 122
 QY 257 AQORVLRVYGFYFGEKLEKEIKLLVCNIDGCLTNGHIYVSGDQKEIISVDVKAIGISLLKK 316
 DB 123 AQORVLRVYGFYFGEKLEKEIKLLVCNIDGCLTNGHIYVSGDQKEIISVDVKAIGISLLKK 182

QY 317 SGIEVELISERACSKOTLSKLDCKMEVSVSKLAVVDWRKEMGLCWKEVAYL 371
 DB 183 SGIEVELISERACSKOTLSKLDCKMEVSVSKLAVVDWRKEMGLCWKEVAYL 237
 RESULT 12
 ADB64227
 ID ADB64227 standard; protein; 202 AA.
 XX AC ADB64227;
 XX DT 04-NOV-2002 (first entry)
 XX DE Human ORF597.
 XX KW Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;
 KW Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;
 KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;
 KW cancer; cardiovascular disease; allergy; autoimmune disease;
 KW wound healing; blood coagulation disorder; inflammatory disorder.
 XX OS Homo sapiens.
 XX PN US2002082206-A1.
 XX PD 27-JUN-2002.
 XX PF 30-MAY-2001; 2001US-00867550.
 XX PR 30-MAY-2000; 2000US-0208427P.
 XX PA (LEAC/) LEACH M D.
 XX PA (MEHR/) MEHRABAN F.
 XX PA (CONL/) CONLEY P B.
 XX PA (TOPP/) TOPPER J N.
 XX PA (LAWD/) LAW D.
 XX PI Leach MD, Mehraban F, Conley PB, Topper JN, Law D;
 XX WPI; 2002-626554/67.
 XX N-PSDB; ABQ98790.
 XX New polypeptide designated ORFX are present in human atherogenic cells
 and are useful to prevent and treat CRFX-associated disorders including
 cancer, allergy, wound healing or autoimmune, cardiovascular or
 inflammatory disease.
 XX Claim 10; SEQ ID NO 1194; 78pp; English.
 XX The present invention relates to novel human ORFX polypeptides and their
 coding sequences (ABP64681 and ABQ98194-ABQ99267). The sequences
 were discovered in human atherogenic cells, in particular in platelets
 and human umbilical vein endothelial cells (HUVEC) and are expressed in
 many other tissues as well. Atherogenic cells are cells which have the
 potential to develop atherosclerotic plaques. The ORFX polypeptides and
 nucleic acids are useful for treating or preventing a pathological
 condition associated with an ORFX-associated disorder, e.g. cancer,
 cardiovascular disease, allergy, autoimmune disease, wound healing, blood
 coagulation disorders or inflammatory disorders. Note: The sequence data
 for this patent did not form part of the printed specification, but was
 obtained in electronic format directly from the USPTC web site at
 seqdata.uspto.gov/sequence.html?DocID=20020082206
 XX SQ Sequence 202 AA;
 Query Match 41.7%; Score 945; DB 5; Length 202;
 Best Local Similarity 92.2%; Pred. No. 1.6e-88;
 Matches 188; Conservative 1; Mismatches 13; Indels 2; Gaps 1;
 QY 1 MDSVEKGAATSVNPRGRPSRGPPKIQNRSGGQGRGVEKPPHAAALILARSGSKIPL 60
 DB 1 MDSVEKGAATSVNPRGRPSRGPPKIQNRSGGQGRGVEKPPHAAALILARSGSKIPL 58

CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
 CC and antiinfective activity, and may also be involved in the determination
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,
 CC nucleic acids and antibodies may be used in the treatment of cancers,
 CC other proliferative disorders such as psoriasis and benign tumours,
 CC neurological disorders such as epilepsy and Alzheimer's disease,
 CC cardiovascular diseases, immune system disorders, disorders related to
 CC organ transplantation, disorders of tissue growth and regeneration,
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol
 CC storage disease, and infectious diseases caused by viral, bacterial,
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a
 CC source of primers and probes in the detection of ORFX genomic sequences
 CC or transcripts, in the identification and cloning of homologous
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases
 XX
 SQ Sequence 74 AA;

Query Match 15.0%; Score 339; DB 5; Length 74;
 Best Local Similarity 86.3%; Pred. No. 9.8e-27;
 Matches 63; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
 QY 204 RQDWGELYENGSYFAKRLHLEMGYLQGGKWHHTKCELEHSDVDIDWPAEQRVLR 263
 DB 1 RQDWGELYENGSYFAKRLHLEMGYLQGGKWHHTKCELEHSDVDIDWPAEQRVLR 60
 QY 264 YGYFGKELKEIK 276
 DB 61 FGYFGKELKEIK 73

RESULT 15
 AAM16058
 ID AAM16058 standard; protein; 44 AA.
 XX
 AC AAM16058;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Peptide #2492 encoded by probe for measuring cervical gene expression.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-05000670.
 XX
 PP 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human cervical epithelial cells.
 XX

PS Claim 27; SEQ ID NO 20884; 487pp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENP; see AAM16058-AA128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 44 AA;

Query Match 10.6%; Score 241; DB 4; Length 44;
 Best Local Similarity 100.0%; Pred. No. 5.4e-17;
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 188 REVTEPLNLNPAKRRPRQDWGELYENGSYFAKRLHLEMGYLQ 231
 DB 1 REVTEPLNLNPAKRRPRQDWGELYENGSYFAKRLHLEMGYLQ 44

Search completed: May 6, 2004, 09:04:19
 Job time : 67.1848 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:00:50 ; Search time 16.485 Seconds
(without alignments)
2532.427 Million cell updates/sec

Title: US-09-930-440B-4
Perfect score: 2266
Sequence: 1 MDSVEKGAATSVSNPRGRPS.....REFAEHICLLMEKYNNSCQK 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: Pir1.*
- 2: Pir2.*
- 3: Pir3.*
- 4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	230.5	10.2	163	2 D70486	conserved hypotet
2	217	9.6	228	2 B53384	polysialic acid ca
3	210.5	9.3	187	2 AC0435	conserved hypotet
4	209.5	9.2	179	2 B83087	conserved hypotet
5	205	9.0	178	2 C81785	conserved hypotet
6	202.5	8.9	395	2 S70912	CMP-N-acetylneuram
7	201.5	8.9	188	2 H65110	hypotetrical 20.0
8	201.5	8.9	188	2 S91138	hypotetrical prote
9	201.5	8.9	188	2 H85983	hypotetrical prote
10	200	8.8	164	2 B64716	conserved hypotet
11	199.5	8.8	164	2 G71802	hypotetrical prote
12	199.5	8.8	188	2 AC0905	conserved hypotet
13	194	8.6	178	2 H81207	hypotetrical prote
14	193.5	8.5	180	2 G64174	hypotetrical prote
15	189.5	8.4	419	2 A36509	N-acylneuraminat
16	185.5	8.2	185	2 B82065	conserved hypotet
17	185.5	8.2	228	2 C64114	CMP-NeuNAc synthet
18	183	8.1	182	2 A82684	conserved hypotet
19	172.5	7.6	235	2 H81276	N-acylneuraminat
20	171	7.5	235	2 S70695	ptmB protein - Cam
21	167.5	7.4	232	2 D81274	N-acylneuraminat
22	165	7.3	162	2 F81413	hypotetrical prote
23	153.5	6.8	517	2 F64560	CMP-N-acetylneuram
24	145	6.4	536	2 B81319	N-acylneuraminat
25	131.5	5.8	246	2 C48492	3-deoxy-manno-octu
26	131	5.8	254	1 G64045	3-deoxy-manno-octu
27	128	5.6	225	2 S75498	hypotetrical prote
28	119.5	5.3	248	2 A26322	3-deoxy-manno-octu
29	119.5	5.3	250	2 AB0171	probable 3-deoxy-m

30 118.5 5.2 248 2 A99754
31 118.5 5.2 248 2 G85617
32 117.5 5.2 228 2 S75394
33 116.5 5.1 248 2 AB0615
34 113.5 5.0 234 2 F70360
35 110.5 4.9 2194 1 JQ1977
36 109 4.8 592 2 D88712
37 108.5 4.8 491 2 A69123
38 108 4.8 239 2 H72667
39 107.5 4.7 502 2 S38139
40 106.5 4.7 722 2 T20362
41 106 4.7 634 2 S77096
42 105.5 4.7 246 2 D97765
43 104 4.6 327 2 C69419
44 101.5 4.5 252 2 A82146
45 99 4.4 475 2 A34443

ALIGNMENTS

RESULT 1
D70486
conserved hypothetical protein aq_2171 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 23-Dec-2002
C:Accession: D70486
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; C
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: D70486
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-163 <AOF>
A:Cross-references: GB:AB000776; NID:G2984355; PIDN:AA007894.1; PID:G2984375; GB:AE0006
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_2171
C:Superfamily: phosphatase

Query Match 10.2%; Score 230.5; DB 2; Length 163;
Best Local Similarity 35.0%; Pred. No. 1.3e-10;
Matches 57; Conservative 39; Mismatches 54; Indels 13; Gaps 3;
QY 269 KEKLEIKLLVCNIDGCLTNGHIYVSGPKELIISYDVXDAIGISLLKSGIEVRLISERA 328
DB 4 RDRVKLLIMDIDGVLTDGKLYTTEGETIKVFNVDGIGIKLQKNGITLAVISGRD 63
QY 329 CSKQTLTKLDCRMVSVSD-----KLAVDWMKEMGLCWKRVAYLGNVSVDEEL 381
DB 64 -SAPLITRLK-----ELGVEIYTGVSVKLIEYKIKESYLSKDEEIGFGEDVVDEWM 117
QY 382 KVGLSGAPADACSYAQKAVGYICKNGCGCAIRFAEHICLL 424
DB 118 KVGFPFVAVRNVAEVRKAVYITQRNGGEGALREVAELIHL 160

RESULT 2
B53384
Polysialic acid capsule biosynthesis protein SiaB NMB0069 [imported] - Neisseria mening
C:Species: Neisseria meningitidis
C>Date: 25-Aug-1995 #sequence_revision 25-Aug-1995 #text_change 19-Jan-2001
C:Accession: B53384; S60759; C61241
R:Swartley, J.S.; Stephens, D.S.
J. Bacteriol. 176, 1530-1534, 1994
A:Title: Identification of a genetic locus involved in the biosynthesis of N-acetyl-D-m
A:Reference number: A53384; MUID:94156865; PMID:8113198
A:Accession: B53384
A:Molecule type: DNA
A:Residues: 1-228 <SWA>

RESULT 4
583087
conserved hypothetical protein PA4458 [imported] - *Pseudomonas aeruginosa* (strain PAO1)
C/Species: *Pseudomonas aeruginosa*
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 23-Dec-2002
C/Accession: B83087
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Eick-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Jørgen, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic patho-
gen
A/Reference number: B82950; MUID:20437337; PMID:10394043
A/Accession: B83057

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-179 <STO>
A;Cross-references: GB:AE004360; GB:AE004091; NID:G9950692; PIDN:AAG07846.1; GSPDB:GN0
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA4458
C;Superfamily: phosphatase

RESULT 5
C81785
conserved hypothetical protein NMA2134 [imported] - Neisseria meningitidis (strain Z2491)
C/Species: Neisseria meningitidis
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 23-Dec-2002
C/Accession: C81785
R/Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More
S.; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandrea
Nature 404, 502-506, 2000
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A/Reference number: A81775; MUID:20225556; PMID:10761919

Query Match: 9.0% Score 205; DB 2; Length 178;
Best Local Similarity 32.9%; Pred. No. 1.4e-08;
Matches 5% Conservative 35; Mismatches 63;
Indels 6; Gaps 3;

C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 23-Dec-2002
 C:Accession: H85983
 R:Perera, N.T.; Flunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 469, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; PMID:12074935; PMID:11206551
 A:Accession: H85983
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-188 <STO>
 A:Cross-references: GB:AE005174; NID:g12517816; PIDN:AA058332.1; GSPDB:GN00145; UNQCP:Z49
 A:Experimental source: strain O157:H7, substrain EDL333
 C:Genetics:
 A:Gene: yrbI
 C:Superfamily: phosphatase

Query Match 8.9%; Score 201.5; DB 2; Length 188;
 Best Local Similarity 29.2%; Pred. No. 2.8e-08;
 Matches 49; Conservative 42; Mismatches 64; Indels 13; Gaps 2;
 QY 271 KLEIKLVNIDGCLTNGHIYVSGDQKEIIISYVDKDAIGISLLKSGIEVRLIS----- 325
 DB 22 KAENIRLLILDVGLSDGLIYMGNGEELKAFNRDVGIRCALTSIEVAIIITGRKAK 81
 QY 326 --ERACSKQTLSSLLDKCKMEVSVDKLAVDWEKEMGLCWKEVAYLNGEVSDEECLR 383
 DB 62 LVEDRCA--TLGIVHL-----YQSQSNKLIATFADLLEKTAIPENVAYGDDLDLMPVMEK 135
 QY 384 VLSGAPADACSYAKVAGYICKNGRGGAIRFAEHICLLMEKVNNS 431
 DB 136 VLSVAVADAHLPLPRADYVTRIAGRGAVREVCVCLLLLAQKIDEA 183

RESULT 12
 B64716
 conserved hypothetical protein HP1570 - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 23-Dec-2002
 C:Accession: B64716
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, E.
 Nature 388, 539-547, 1997
 A:Authors: Mallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; PMID:97394467; PMID:9252185
 A:Accession: B64716
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-154 <TOM>
 A:Cross-references: GB:AE000654; GB:AE000511; NID:g2314743; PIDN:AAD08607.1; PID:g231475
 C:Superfamily: phosphatase

Query Match 8.8%; Score 200; DB 2; Length 164;
 Best Local Similarity 29.2%; Pred. No. 3.1e-08;
 Matches 49; Conservative 46; Mismatches 57; Indels 16; Gaps 3;
 QY 275 IKLLVNCIDGCLTNGHIYVSGDQKEIIISYVDKDAIGISLLKSGIEVRLISERACSKQTL 334
 DB 2 IKLLLDVDTGTLTDSGLYFDENPHEIKAFNRDVGIRCALTSIEVAIIITGR----- 54
 QY 335 SSLKLDCKME-----VSVDKLAVDWEKEMGLCWKEVAYLNGEVSDEECLRVCL 386
 DB 55 TSIMVKRMESLGVQFVFNVENKNTVIERLKKDLQLSAQEIAACVGDYNDLGMFKACAL 114
 QY 387 SGAPADACSYAKVAGYICKNGRGGAIRFAEHICLLMEKVNNS 434
 DB 115 SPAPFDAHPLLKSKAYKVLQNSGGKAVREADYLTLEGLQCEALX 161

RESULT 11

G71802
 hypothetical protein jhp1478 - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Dec-2002
 C:Accession: G71802
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat
 A:Reference number: A71800; PMID:99120557; PMID:9923682
 A:Accession: G71802
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-164 <ARN>
 A:Cross-references: GB:AE001569; GB:AE001439; NID:g4156095; PIDN:AAD07050.1; PID:g41560
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp1478
 C:Superfamily: phosphatase

Query Match 8.8%; Score 199.5; DB 2; Length 164;
 Best Local Similarity 30.3%; Pred. No. 3.4e-08;
 Matches 47; Conservative 43; Mismatches 50; Indels 15; Gaps 2;
 QY 275 IKLLVNCIDGCLTNGHIYVSGDQKEIIISYVDKDAIGISLLKSGIEVRLISERACSKQTL 334
 DB 2 IKLLLDVDTGTLTDSGLYFDENPHEIKAFNRDVGIRCALTSIEVAIIITGR----- 54
 QY 335 SSLKLDCKME-----VSVDKLAVDWEKEMGLCWKEVAYLNGEVSDEECLRWGL 386
 DB 55 TSIMVKRMESLGVQFVFNVENKNTVIERLKKDLQLSAQEIAACVGDYNDLGMFKACTL 114
 QY 387 SGAPADACSYAKVAGYICKNGRGGAIRFAEHI 421
 DB 115 SPAPFDAHPLLKSKAYKVLQNSGGKAVREADYL 149

RESULT 12
 AC0905
 conserved hypothetical protein SRY3495 [imported] - Salmonella enterica subsp. enterica
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 23-Dec-2002
 C:Accession: AC0905
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
 Th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serc
 A:Reference number: AB0502; PMID:12153497; PMID:11677608
 A:Accession: AC0905
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-188 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD07833.1; PID:g16504381; GSPDB:GX00176
 C:Genetics:
 A:Gene: SRY3495
 C:Superfamily: phosphatase

Query Match 8.8%; Score 199.5; DB 2; Length 188;
 Best Local Similarity 30.4%; Pred. No. 4e-08;
 Matches 51; Conservative 40; Mismatches 64; Indels 13; Gaps 3;
 QY 271 KLEIKLVNIDGCLTNGHIYVSGDQKEIIISYVDKDAIGISLLKSGIEVRLIS----- 325
 DB 22 KAENIRLLILDVGLSDGLIYMGNGEELKAFNRDVGIRCALTSIEVAIIITGRKAK 81
 QY 326 --ERACSKQTLSSLLDKCKMEVSVDKLAVDWEKEMGLCWKEVAYLNGEVSDEECLR 383
 DB 82 LVEDRCA--TLGIVHL-----YQSQSNKLIATFADLLEKTAIPENVAYGDDLDLMPVMEK 135

QY 384 VGLSGAPADACSYAKAVGYICKNGGSGATREPAEHICLIMKXNNS 431
 Db 136 VGLSVAVADAPHLIPRXYTHIAGRGAVREVCDLLLAQGLDEA 183

RESULT 13
 H61207
 conserved hypothetical protein NMB0353 (imported) - Neisseria meningitidis (strain MC58)
 C:Species: Neisseria meningitidis
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 23-Dec-2002
 C:Accession: H61207
 R:Tectelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, C.; Fleischmann, R.D.; Dougherty, B.A.;
 Xi, K.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Eizsa, M.
 Science 237, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A81000; MUID:20175755; PMID:10711037
 A:Accession: H61207
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-178 <TET>
 A:Cross-references: GB:AE002392; GB:AE002098; NID:G7225573; PIDN:AAE40796.1; PID:G722557
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB0353
 C:Superfamily: phosphatase

Query Match 8.6%; Score 194; DB 2; Length 178;
 Best Local Similarity 31.6%; Pred. No. 1e-07;
 Matches 49; Conservative 36; Mismatches 64; Indels 6; Gaps 3;

QY 274 EIKLLVNCIDCLTNGHYVSGDQKEILSYDVVDKDAIGLSLLKSGIEVRLISERACSKQT 333
 Db 14 KIKLLLDVPSVLDGFRIFRDNGEELKSFETLDGHLKMLQASGVQTAITERDPSVG 73

QY 334 LSSLLKLDCKMEV-SVSKLAVVDWRKEMGLCWKEVAYLGNVEVDEECLKEVLSGAPAD 392
 Db 74 IRVQLGINYFKGISDKRAYEELRACAGVEEACAFVGDVVDLPVWRCGLPVAVPG 133

QY 393 ACSYAKAVGYICKNGGSGATREPAEHIC-LME 426
 Db 134 AHWFTQRAAYITEHAGGAGVRE---VCDLIWQ 164

RESULT 14
 G64174
 hypothetical protein H11679 - Haemophilus influenzae (strain Rd KW20)
 C:Species: Haemophilus influenzae
 C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 23-Dec-2002
 C:Accession: G64174
 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
 Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Weidman, J.
 D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64030; MUID:95350630; PMID:7542800
 A:Accession: G64174
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-180 <TIGR>
 A:Cross-references: GB:U32841; GB:LA2023; NID:gl574529; PIDN:AAE23325.1; PID:gl574531; T
 A:Note: best homolog was a hypothetical protein from Escherichia coli
 C:Superfamily: phosphatase

Query Match 8.5%; Score 193.5; DB 2; Length 180;
 Best Local Similarity 29.0%; Pred. No. 1.1e-07;
 Matches 45; Conservative 41; Mismatches 66; Indels 3; Gaps 2;

QY 269 KEKLEIKLLVNCIDCLTNGHYVSGDQKEILSYDVVDKDAIGLSLLKSGIEVRLISER- 327
 Db 2 QQKLENIKFTVDVGVLDGVLTDQLHYDANGBAIKSFHVRDLGLIKMLMDADIQAVVLSGRD 61

QY 328 -ACSKQTLSLLKLDCKMEVSVSKLAVVDWRKEMGLCWKEVAYLGNVEVDEECLKRVGL 386
 Db 62 SPILRRRIADLGKLFLLGKLEKETACFD-LMKQAGVTAEQTAYIGDDSDVLPAPFAACGT 120

QY 387 SGAPADACSYAKAVGYICKNGGSGATREPAEHI 421
 Db 121 SFAVADAPYVKNADVHVLSTHGKGRFEMSDMI 155

RESULT 15
 A36509
 N-acetylneuraminase cytidyltransferase (EC 2.7.7.43) - Escherichia coli
 C:Species: Escherichia coli
 C>Date: 01-Feb-1991 #sequence_revision 01-Feb-1991 #text_change 08-Oct-1999
 C:Accession: A36509
 J:Zapata, G.; Vann, W.F.; Arcanson, W.; Lewis, M.S.; Moos, M.
 J. Biol. Chem. 264, 14769-14774, 1989
 A:Title: Sequence of the cloned Escherichia coli K1 CMP-N-acetylneuraminic acid synthet.
 A:Reference number: A36509; MUID:89359273; PMID:2549035
 A:Accession: A36509
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-419 <ZAP>
 A:Cross-references: GB:J05023; NID:gl46943; PIDN:AAE24210.1; PID:gl46944
 C:Keywords: nucleotidyltransferase

Query Match 8.4%; Score 189.5; DB 2; Length 419;
 Best Local Similarity 22.4%; Pred. No. 7e-07;
 Matches 88; Conservative 66; Mismatches 151; Indels 87; Gaps 11;

QY 45 LAAILARGGSKGPLKNIKHLAGVPLIGWVRALDSGAFQSVVSTHDEIENVAQKF 104
 Db 5 IIAIIPARSGSKGRNKNALMLIDKPLIAYTIEALQSEMFEKVIITDSEYGAIESY 64

QY 105 GAQVHRSSSVSKDSTSLDAIIEFLVYXNEXDIVGNTQATXCLHPTDLOKVAEWIRE- 163
 Db 65 GADPELLPEELATDKASFEFKHALSIYDYFSFALLQPTSPFPDSTHIEAVKLYOTL 124

QY 164 EGYDSXFSVVRHQRFRWSEIQGVREVTPL-----NLNPAKRPRQWWDGELYEN 214
 Db 125 EKYQCVSVVTRSNK-----PSQIRPLDDYSTLSFFDLDSKYNNNS--IVEYHEN 173

QY 215 GSFYFA-KRHLEMGYLOGGKWHHTKCELSHSVDIVDIDWPIA---EQRVLRGYGFCKE 270
 Db 174 GAIFIAKQHYLHTKFFGRYSLAYIMDKSSSLDIDRMDPELAITIQO-----KK 224

QY 271 KLKEIKLLVNCIDCLTNGHYVSGDQKEILSYDVVDKDAIGLSL-----KKSGETETLI 324
 Db 225 NRQKIDL-----YONIHNRINERNEFDVSVDITLIGHSLFDYMDVKKINDIEVNNL 276

QY 325 SERACSKQTLSLLKLDCKMEVSVSKL-----AVVDWRKEMGLCWKEVAYLGNVEV 376
 Db 277 GIAGINSKEYVEYIIEKELIVNFGFVFIIFGTNDIVVSDWKE----- 320

QY 377 DEECLKRVGLSGAPADACSYAKAVGYICKN 408
 Db 321 -----DTLWYLLKKTQYIKKN 337

Search completed: May 6, 2004, 09:09:42
 Job time : 17.735 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM : protein - protein search, using sw model

Run on: May 6, 2004, 08:56:04 ; Search time 11.4127 Seconds
(without alignments)
1980.112 Million cell updates/sec

Title: US-09-930-440B-4

Percent score: 2266

Sequence: 1 MDSVEKGAATSVNPRGRPS.....REFAHICLLMEKVNNSCQK 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Sea thed: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Min num DB seq length: 0

Max num DB seq length: 2000000000

Pos -processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result ID	Score	Query Match %	Length	DB ID	Description
1	217	9.6	228	1 NEUA_NEIMB	Q57385 neisseria m
2	210.5	9.3	187	1 KDSC_YERPE	Q8z347 versinia pe
3	202.5	8.9	413	1 NEUA_STRAS	Q8z398 streptococ
4	201.5	8.9	188	1 KDSC_ECOL6	Q8fd72 escherichia
5	201.5	8.9	188	1 KDSC_ECOL6	P45396 escherichia
6	199.5	8.8	188	1 KDSC_SALTI	Q8z395 salmonella
7	199.5	8.8	188	1 KDSC_SALTI	Q8z180 salmonella
8	198.5	8.8	413	1 NEUA_STRAS	Q8af59 streptococ
9	193.5	8.5	180	1 KDOP_HAEIN	P45314 haemophilus
10	189.5	8.4	419	1 NEUA_ECOLI	P23266 escherichia
11	185.5	8.2	228	1 NEUA_HAEIN	Q57140 haemophilus
12	171	7.5	235	1 PTMB_CAMCO	Q45982 campylobact
13	163.5	7.2	258	1 KDSC_PASMU	P57883 pasteurella
14	131.5	5.8	245	1 KSU5_ECOLI	P42216 escherichia
15	131	5.8	253	1 KDSE_HAEIN	P44490 haemophilus
16	124	5.5	393	1 GLGC_CLOPE	Q8xp97 clostridium
17	121.5	5.4	231	1 KDSE_VIEBU	Q8ada9 vibrio vuln
18	119.5	5.3	247	1 KDSE_ECOLI	P04951 escherichia
19	119.5	5.3	250	1 KDSE_YERPE	Q8z944 versinia pe
20	118.5	5.2	247	1 KDSE_ECO57	Q8xdg6 escherichia
21	116.5	5.1	247	1 KDSE_SALTI	Q8z800 salmonella
22	116.5	5.1	247	1 KDSE_SALTY	Q8zq40 salmonella
23	114.5	5.1	247	1 KDSE_ECOL6	Q8fj49 escherichia
24	113.5	5.0	234	1 KDSE_AQUAE	Q66914 aquifex aeo
25	110.5	4.9	2194	1 GLSN_MEDSA	Q03460 medicago sa
26	109	4.8	251	1 KDSE_VIEBU	Q87x14 vibrio para
27	108.5	4.8	491	1 PYCA_METTH	Q27939 methanobact
28	107.5	4.7	502	1 LAGI_YEAST	P36146 saccharomyc
29	106.5	4.7	722	1 GPM_CAREL	P90795 caenorhabdi
30	106	4.7	634	1 ICFG_SNNY3	P37979 synchocyst
31	105.5	4.7	246	1 KDSE_RICCN	Q92196 rickettsia
32	105	4.6	475	1 NIFB_ANAAZ	Q43883 anabaena az
33	104	4.6	245	1 KDSE_FUSNN	Q8rfa8 fusobacteri

34	101.5	4.5	252	1 KDSB_VIBCH	Q9kpx2 vibrio chol
35	101.5	4.5	760	1 METE_VIBHA	Q8krg6 vibrio harv
36	99.5	4.4	254	1 KDSB_PSRPK	Q881m7 pseudomonas
37	99	4.4	475	1 NIFB_ANASP	P20627 anabaena sp
38	96	4.2	254	1 KDSB_PSESM	Q87yf7 pseudomonas
39	96	4.2	293	1 ERA_MICPU	Q98q11 mycoplasma
40	95.5	4.2	246	1 KDSB_RICPR	Q9zdf0 rickettsia
41	95	4.2	71	1 KSUI_ECOLI	P42215 escherichia
42	94.5	4.2	1076	1 CARB_HALER	Q8rs53 halomonas e
43	94.5	4.2	3084	1 LMAL_MOUSE	P19137 mus musculu
44	94.5	4.2	8797	1 SNEI_HUMAN	Q8sf91 homo sapien
45	93.5	4.1	340	1 Y483_METJA	Q57907 metatarsococ

ALIGNMENTS

RESULT 1
NEUA_NEIMB
ID NEUA_NEIMB STANDARD; PRT; 228 AA.
AC Q57385;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 13-OCT-2003 (Rel. 42, Last annotation update)
DE Acy-neuraminidase cytidyltransferase (EC 2.7.7.43) (CMP-N-
DE acetylneuraminic acid synthetase) (CMP-NeuNAC synthetase) (CMP-sialic
DE acid synthetase).
GN NEUA OR SIAB OR SYNB OR NMB0069.
OS Neisseria meningitidis (serogroup B), and
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491, 487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B1940 / Serogroup B;
RX MEDLINE=93012891; PubMed=1398032;
RA Edwards U., Prosch M.;
RT "Sequence and functional analysis of the cloned Neisseria
RT meningitidis CMP-NeuNAC synthetase.";
RL FEMS Microbiol. Lett. 75:161-166(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NMB / Serogroup B;
RX MEDLINE=94156865; PubMed=8113198;
RA Swartley J.S., Stephens D.S.;
RT "Identification of a genetic locus involved in the biosynthesis of N-
RT acetyl-D-mannosamine, a precursor of the (alpha 2-->8)-linked
RT polysialic acid capsule of serogroup B Neisseria meningitidis.";
RN [3]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NTC 8249 / Serogroup B;
RX MEDLINE=94321329; PubMed=8045888;
RA Ganguli S., Zapata G., Wallis T., Reid C., Boulnois G.J., Vann W.F.,
RA Roberts I.S.;
RT "Molecular cloning and analysis of genes for sialic acid synthesis in
RT Neisseria meningitidis group B and purification of the meningococcal
RT CMP-NeuNAC synthetase enzyme.";
RL J. Bacteriol. 176:4583-4583(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=NRCC 4030 / 406Y / Serogroup Y;
RA Gilbert M., Watson D.C., Wakarchuk W.W.;
RT "Purification and characterization of the recombinant CMP-sialic acid
RT synthetase from Neisseria meningitidis.";
RL Biotechnol. Lett. 19:417-420(1997).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Feden J.F., Dodson R.J.,

```

RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mascen T., Ciecko A., Parksey D.S., Blair E., Clitstone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzia M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC -!- CATALYTIC ACTIVITY: CTP + N-acetylneuraminate = diphosphate + CMP-N-
CC acylneuraminate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the CMP-NeuNAc synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M95053; AAA20476.1; -
DR EMBL; U04328; AAA17655.1; -
DR EMBL; X78068; AAA54983.1; -
DR EMBL; U60146; AAB60780.1; -
DR EMBL; AE002366; AAF40536.1; -
DR PIR; B53384; B53384.
DR PDB; 1EYR; 14-FEB-01.
DR PDB; 1E2I; 14-FEB-01.
DR TIGR; NMB0069; -.
DR InterPro; IPR003329; Cytidylyl trans.
DR Pfam; PF02348; CTP trans 3; 1.
KW Transferase; Nucleotidyltransferase; Sialic acid; Complete proteome;
KW 3D-structure.
SQ SEQUENCE 228 AA; 24892 MW; F2510733BEC3A31 CRC64;

Query Match 9.6%; Score 217; DB 1; Length 228;
Best Local Similarity 27.2%; Pred. No. 5.6e-10;
Matches 65; Conservative 40; Mismatches 82; Indels 52; Gaps 6;

QY 47 ALLIARGSKGIPLKNVXHLGAVPLGVLRALDSGAFQSVVWTDHDEINVAKQFGA 106
D 7 AVILARQNSKGLPLKXLRNMGISLHTTINAISSKCFDRIIVTDGGLIAEAKNFV 66
QY 107 QVRRSSSVKDSSTLDAITFELNXXEDTVGNIQATSKCLEPTDLQKVAEMIRESGY 166
D 67 EVLRLPALASDTASSISGVIAL-----ETIGNSGVTVLLQPTSLRTGAHIRE--- 117
QY 167 DSXFSV-----VRRHQFRWSEIQGVREYVTEPLNLNPAKRPRRQ----- 205
D 118 -AFSLFDEKIGSVSVSACPMVEHPLK-TLQINNGEYAPNKHLSDLBQPRQQLPQAFRP 174
QY 206 -----DWDGELIYENGSYFAKHLIEMGYLGGKWHTKCELSHSDVIDWDPIAE 258
D 175 NGAIYINDTASLIANNCFIAPTFLTYMSH-----QDSIDIDTLELDLQQAEE 220

RESULT 2
KDSC VERPE
ID_KDSC_VERPE STANDARD; PRT; 187 AA.
AC Q8ZB47;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)
DE (KDO 8-P phosphate).
DE GN KDSC OR YPO3578 OR Y0150.
DE Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
CX NCBI_TaxID=632;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Farhikhi J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moute S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-KIMS / Biovar Mediaevalis;
RX MEDLINE=22137853; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
CC -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic
CC phosphate (By similarity).
CC -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate + phosphate +
CC H(2)O = 3-deoxy-D-manno-octulosonate + phosphate.
CC -!- COFACTOR: Magnesium (Probable).
CC -!- PATHWAY: Lipopolysaccharide biosynthesis; KDO biosynthesis; third
CC step.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SIMILARITY: Belongs to the kdsC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ414157; CAC92807.1; -
DR EMBL; AE013615; AAM83744.1; -
DR InterPro; IPR006549; HAD-SF-IIIa.
DR InterPro; IPR005834; Hydrolase.
DR InterPro; IPR008230; Sugar_Ptase.
DR Pfam; PF00702; Hydrolase; 1.
DR PIRSF; PIRSF006118; Sugar_Ptase; 1.
DR TIGRfam; TIGR01662; HAD-SF-IIIa; 1.
KW Lipopolysaccharide biosynthesis; Hydrolase; Magnesium;
KW Complete proteome.
SQ SEQUENCE 187 AA; 20299 MW; 295EA75C7DEA7878 CRC64;

Query Match 9.3%; Score 210.5; DB 1; Length 187;
Best Local Similarity 29.1%; Pred. No. 1.4e-09;
Matches 52; Conservative 43; Mismatches 73; Indels 11; Gaps 3;

QY 255 PIAEQRLRYGVFGKELKEIKLVNCDGLTNGHIYVSGDQKEIISVDKDAIGISLL 314
D 13 PVADDVI-----QRAANIRLLICDVGWSDGLIYNGQGEELKAFNVDGYGIRCL 64
QY 315 KKSGLVRLISERACS--KQTLSIKLDCRKEVSDKLVAVDVEWRKEMGLCWKEVAYIG 372
D 65 ITSDIDVAITIGRAKULEDRANTLGI-THLYQQSDKLVAYHELLATLQCOPEQVAIG 123
QY 373 NEVSDDECLKRVGLSGAPADACSYAKVGVYCKNGGEGAIRFPAEHCILMEKNNS 431
D 124 DDLIDWPWMAQVGLSVAVADAPHLPLPKAHYVTRIKGGRGAVREVCDLILLAQDKLEGA 182

RESULT 3
NEUA STRA3
ID NEUA_STRA3 STANDARD; PRT; 413 AA.

```



```

RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RL flexneri serotype 2a strain 2457T.";
CC Infect. Immun. 71:2775-2786(2003).
CC
CC -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic
CC phosphate (By similarity).
CC -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate 3-phosphate +
CC H2O = 3-deoxy-D-manno-octulosonate + phosphate.
CC -!- COFACTOR: Magnesium (Probable).
CC -!- PATHWAY: Lipopolysaccharide biosynthesis; KDO biosynthesis; third
CC step.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SIMILARITY: Belongs to the kdsC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016767; AAN82388.1; -
CC DR EMBL; AE015335; AAN44704.1; -
CC DR EMBL; AE016989; AAP18518.1; -
CC DR InterPro; IPR006549; HAD-SF-IIIa.
CC DR InterPro; IPR005834; Hydrolase.
CC Pfam; PF00702; Hydrolase; 1.
CC DR TIGRFAMs; TIGR01662; HAD-SF-IIIa; 1.
CC DR Lipopolysaccharide biosynthesis; Hydrolase; Magnesium;
CC Complete proteome.
CC KW
CC SEQUENCE 188 AA; 20015 MW; 6A40ED8C2CB1E3A7 CRC64;
CC
CC Query Match 8.9%; Score 201.5; DB 1; Length 188;
CC Best Local Similarity 29.2%; Pred. No. 7.3e-09;
CC Matches 49; Conservative 42; Mismatches 64; Indels 13; Gaps 2;
CC
CC 271 KLKEIKLVNIDCLTNGHYVSGDOKEILSYDVKDAIGLSLKKSGIEVRLIS-----325
CC 22 KAENIRLLDVGSLDGLIYMGNGEELKAFNVRGIGRCALTSIDIEVAITGRKAK 81
CC
CC 326 --ERACKQTLSSIKLCKMEVSVSKLAVVDEVRKEMGLCKWEKVAIAGNEVSDDECLR 383
CC 82 LVDEKCATLGLITL-----YQGSNKLIAFSDLEKLAIAPEVAVYVGDLDIDWPVWEK 135
CC
CC 384 VGLSGADACSYAQKAVGICKNGGRCATREFAEHCILMKVKNNS 431
CC 136 VGLSVAVADAEHLIPRADYVTRIAGRGAVREVCDLLLAQGLDEA 183
CC
CC RESULT 5
CC KDS_C_ECOLI STANDARD; PRT; 188 AA.
CC
CC AC P45396; P45398;
CC DT 01-NOV-1995 (Rel. 32, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)
CC DE (KDO 8-P phosphatase).
CC GN KDS_C OR B3198 OR Z4561 OR ECS4077.
CC OS Escherichia coli, and
CC OS Escherichia coli C157:H7.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC OC Enterobacteriaceae; Escherichia.
CC OX NCBI_TaxID:562, 83334;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=K12 / MG1655;
CC RX MEDLINE=97426617; PubMed=9278053;
CC RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
CC Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
CC Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

```

Query Match:

8.9%; Score 201.5; DB 1; Length 188;

Best Local Similarity 29.2%; Pred. No. 7.3e-09;
Matches 49; Conservative 42; Mismatches 64; Indels 13; Gaps 2;

QY 271 KKEIKLLVCNIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKSGIEVRLIS----- 325
DB 22 KAENIRLLILDVGVLSGLIYMGNGEELKAFNRDGYGIRCALTSNIEVAITGRKAK 81

QY 326 --ERACKTSLKLDCKMEVSVSKLAVVDWRKEMGLCWKEVAYLNGVSEDECKR 383
DB 82 LVEDRCATLGIHTL-----YQGSNKLIAFSLLEKLAIPENVAIVGDDLDWPMVK 135

QY 384 VGLSGAPADACSVAKAVGICKNCGRGRAIRFAEHICLLMEKVNNS 431
DB 136 VGLSVAADAHPLIIPRADIYVTHIAGRGAVREVCOLLIAQGLDEA 183

RES 1T 6
KDS SCALTI
ID KDS SCALTI STANDARD; PRT; 188 AA.

AC Q823G5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)
DE (KDO 8-P phosphatase)
GN KDS OR STY3495 OR T3233.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham P., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krost A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RE [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic phosphate (By similarity).
CC -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate 8-phosphate + H2O = 3-deoxy-D-manno-octulosonate + phosphate.
CC -!- COFACTOR: Magnesium (Probable).
CC -!- PATHWAY: Lipopolysaccharide biosynthesis; KDO biosynthesis; third step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: Belongs to the kdsC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AL627278; CAD07833.1; -.

DR EMBL; AE016845; AA070769.1; -.
DR InterPro; IPR006549; HAD-SF-IIIa.
DR InterPro; IPR005834; Hydrolase.
DR InterPro; IPR008230; Sugar Phase.
DR Pfam; PF00702; Hydrolase; I.
DR PIRSF; PIRSF006118; Sugar Phase; 1.
DR TIGRFAMS; TIGRF01662; HAD-SF-IIIa; 1.
KW Lipopolysaccharide biosynthesis; Hydrolase; Magnesium;
KW Complete proteome.
SQ SEQUENCE 188 AA; 20075 MW; ADEFB0323DBFEF43 CRC64;

Query Match 3.8%; Score 199.5; DB 1; Length 188;
Best Local Similarity 30.4%; Pred. No. 1e-08;
Matches 51; Conservative 40; Mismatches 64; Indels 13; Gaps 3;

QY 271 KKEIKLLVCNIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKSGIEVRLIS----- 325
DB 22 KAENIRLLILDVGVLSGLIYMGNGEELKAFNRDGYGIRCALTSNIEVAITGRKAK 81

QY 326 --ERACKTSLKLDCKMEVSVSKLAVVDWRKEMGLCWKEVAYLNGVSEDECKR 383
DB 82 LVEDRCATLGIHTL-----YQGSNKLIAFSLLEKLAIPENVAIVGDDLDWPMVK 135

QY 384 VGLSGAPADACSVAKAVGICKNCGRGRAIRFAEHICLLMEKVNNS 431
DB 136 VGLSVAADAHPLIIPRADIYVTHIAGRGAVREVCOLLIAQGLDEA 183

RESULT 7
KDS SCALTI
ID KDS SCALTI STANDARD; PRT; 188 AA.

AC Q82LS0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)
DE (KDO 8-P phosphatase).
GN KDS OR STM3316.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=LT2 / SGSC412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du R., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan K., Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic phosphate (By similarity).
CC -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate 8-phosphate + H2O = 3-deoxy-D-manno-octulosonate + phosphate.
CC -!- COFACTOR: Magnesium (Probable).
CC -!- PATHWAY: Lipopolysaccharide biosynthesis; KDO biosynthesis; third step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: Belongs to the kdsC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE008852; AAL22185.1; -.


```

RN  [2]
RP  FUNCTION, AND COFACTOR.
RX  MEDLINE=22625721; PubMed=12639950;
RA  Wu J., Woodard R.W.;
RT  "Escherichia coli YrbI is 3-deoxy-D-manno-octulosonate 8-phosphate
RL  phosphate."
RN  J. Biol. Chem. 278:18117-18123 (2003).
RP  [3]
RX  X-RAY CRYSTALLOGRAPHY (1.67 ANGSTROMS), AND METAL BINDING.
RX  MEDLINE=21824378; PubMed=11935514;
RA  Parsons J.P., Lim K., Tempczyk A., Krajewski W., Eisenstein E.,
RT  Herzberg O.;
RL  (Hil1679) is a phosphatase."
CC  [From structure to function: YrbI from Haemophilus influenzae
CC  proteins 46:393-404 (2002).]
CC  -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic
CC  phosphate.
CC  -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate 8-phosphate +
CC  H(2)O = 3-deoxy-D-manno-octulosonate + phosphate.
CC  -!- COFACTOR: Magnesium (Probable).
CC  -!- MISCELLANEOUS: Cobalt was used in the crystallography experiment
CC  but magnesium is likely to be the physiological metal.
CC  -!- SIMILARITY: Belongs to the kdsC family.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
CC  EMBL; U32841; AAC23325.1; -.
DR  PIR; G64174; G64174.
DR  PDB; 1J8D; 12-AUG-03.
DR  PDB; 1KIE; 12-AUG-03.
DR  TIGR; Hil1679; -.
DR  InterPro; IPR006549; HAD-SF-IIIa.
DR  InterPro; IPR005834; Hydrolase.
DR  InterPro; IPR008230; Sugar Phase.
DR  Pfam; PF00702; Hydrolase; 1.
DR  PIRSF; PIRSF06118; Sugar Phase; 1.
DR  TIGSFAMS; TIGR01662; HAD-SF-IIIa; 1.
DR  Hydrolase; Metal-binding; Magnesium; 3D-structure; Complete proteome.
FT  METAL 14 14 MAGNESIUM (PROBABLE).
FT  METAL 16 16 MAGNESIUM (PROBABLE).
FT  METAL 107 107 MAGNESIUM (PROBABLE).
SQ  SEQUENCE 180 AA; 19432 MW; 23CD435E4E83A095 CRC64;

Query Match      8.5%; Score 193.5; DB 1; Length 180;
Best Local Similarity 29.0%; Pred. No. 2.9e-08;
Matches 45; Conservative 41; Mismatches 66; Indels 3; Gaps 2;

QY  269 KEKKEIKLVNIGCLTNGHYVSGDQKEIISYVDKDAIGISLILKSGIEVRLISER- 327
DB  2 OOKLNRIKVFITDVGWLTGQLHYDANGEARKSFHRLGDKMLMDADIQVAVLSGRD 61
QY  328 -ACSKQTLSSILKDCMEVSDKLAVDDEKEMGLCWKEVAYLGNESVDEECLKRVGL 386
DB  62 SPILRRRTADIGIKLFFLGKLEKETACFD-LNKQGVTRAEQTAYIGDSVDLPAFAACGT 120
QY  387 SGAPADACSQAQKAVGYICKNGGKGATREFAEHI 421
DB  121 SPVADAPTYVKNVDVHLSVTHGGKGAFFEMSDMI 155

RESULT 10
NEUF ECOLI
ID NEUA ECOLI
AC P13266;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 3e, Last annotation update)

```

```

DE  Acylneuraminate cytidyltransferase (EC 2.7.7.43) (CMP-N-
DE  acetylneuraminic acid synthetase) (CMP-NeuNAc synthetase).
GN  NEUA.
OS  Escherichia coli.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Escherichia.
OX  NCEI_TaxID=562;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=KL;
RX  MEDLINE=89359273; PubMed=2549035;
RA  Zapata G., Vann W.F., Aaronson W., Lewis M.S., Moos M.;
RT  "Sequence of the cloned Escherichia coli Ki CMP-N-acetylneuraminic
RL  acid synthetase gene."
CC  J. Biol. Chem. 264:14769-14774 (1989).
CC  -!- CATALYTIC ACTIVITY: CTP + N-acylneuraminate = diphosphate + CMP-N-
CC  acylneuraminate.
CC  -!- COFACTOR: Magnesium; other divalent cations are less effective.
CC  -!- SUBUNIT: Homodimer (Probable).
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC  -!- SIMILARITY: Belongs to the CMP-NeuNAc synthetase family.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
CC  EMBL; J05023; AAA24210.1; -.
DR  PIR; A36509; A36509.
DR  InterPro; IPR003329; Cytidylyl trans.
DR  InterPro; IPR001087; Lipase GDSL.
DR  Pfam; PF02348; CTP transf_3; 1.
DR  Pfam; PF00857; Lipase GDSL; 1.
DR  Transferase; Nucleotidyltransferase; Sialic acid; Magnesium.
SQ  SEQUENCE 419 AA; 48736 MW; AB8E7783CDD74F19 CRC64;

Query Match      8.4%; Score 189.5; DB 1; Length 419;
Best Local Similarity 22.4%; Pred. No. 1.9e-07;
Matches 88; Conservative 66; Mismatches 151; Indels 87; Gaps 11;

QY  45 LAALILARGSGKIPLNKIKHLAGVPLIGWVLAALDGAQFQSVWVSTHDEIENVAQKF 104
DB  5 IIAIIPARGSGKLRNKNALMLIDKPLLAYTTEALQSEMFEKVIIVTDSQEQGALAESY 64
QY  105 GQVHRSEVSKDSTSLDAIEFLNLYNEXDIVGNIQATSCXCLHPTDLQKVAEMIRE- 163
DB  65 GADELRLPELATDKASSPEFIKHALSIYTDYESFALLOFTSPFRDSTHIEAVKLYQTL 124
QY  164 EGYDSKFSVVRHQFQFWSBIQGVREVTPL-----NLNPAKRPRQDWDGELYEN 214
DB  125 EKYQCQVSVTRNK-----PSQIIPLDYDSTLSPFDLDYSKYNNRS--IVYHPN 173
QY  215 GSFYTA-KRHLEMGYLGQGGKWHHTKCELEHSDVDIDVDIPFA---EQRVLRYGVGKE 270
DB  174 GAFLANKQHYLHTKHFGRYSLAYIMDKESLIDIDRMDFELAITQQ-----KK 224
QY  271 KLEIKLLVCNIDGLTNGHYVSGDQKEIISYVDKDAIGISL-----KKSIEVRLI 324
DB  225 NFQKIDL-----YQNHNRINRKNRNFDSVSDITLGHSLFDYWDVKINDIEVNNL 276
QY  325 SERACKQTLSSILKDCMEVSDKL-----AVDEWRKEMGLCWKEVAYLGNESV 376
DB  277 GIAGINSKEYEYIIIEKELIVNFGFVFIFFGTNDIVSDMKK----- 320
QY  377 DEECLKRVLSGAPADACSQAQKAVGYICKN 428
DB  321 -----DTLWYLNKTKQYIKKN 337

RESULT 11

```

NEUA HAEIN
ID NEUA HAEIN STANDARD; PRT; 228 AA.
AC Q57140; OC5051;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable acylneuraminase cytidyltransferase (EC 2.7.7.43) (CMP-N-acetylneuraminic acid synthetase) (CMP-NeuNAc synthetase) (CMP-sialic acid synthetase).
DN NEUA OR H1279.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
CX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann A.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Karpavich A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips R.A., Spriggs T., Hedblom B., Cotton M.D., Uterback T.R., Hanna M.C., Nguyen D.C., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512 (1995).
CC -!- CATALYTIC ACTIVITY: CMP + N-acetylneuraminic acid = diphosphate + CMP-N-acetylneuraminic acid.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the CMP-NeuNAc synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; U32807; AAC22927.1; --
CC TIGR; H1279;
CC InterPro; IPR003329; Cytidylyl_trans.
CC Pfam; PF02348; CTP_transf_3; 1.
CC Transfase; Nucleotidyltransferase; Sialic acid; Complete proteome.
SQ SEQUENCE 228 AA; 25308 MW; B3CC7674CA379551 CRC64;
Query Match 8.2%; Score 185.5; DB 1; Length 228;
Best Local Similarity 27.9%; Pred. No. 1.7e-07;
Matches 65; Conservative 40; Mismatches 99; Indels 35; Gaps 9;
QY 47 ALIARGSGKGIPLKNIKHLGVLIGWVLRALDAGAFOSVWVSTHDEIENKAKQF 106
Db 9 AIIIPARAGSGKGIPLKNIKHLGVLIGWVLRALDAGAFOSVWVSTHDEIENKAKQF 68
QY 107 QVHRSEVSKDSSTSLDAIEFLNXXEXDIVGNIGTSCXCLHPT-----DLQKVAE 159
Db 69 KPVARPESSAQSDTRTIDAILHCLSTLIS-----QGTAAALQPTSPRLNALDIRNAME 122
QY 160 MIREBGYSXFSV--VRHQFF-----WSEIQGVRETEPLNINPAKRPRQDWDGELY 212
Db 123 IFLGKYKSVSVSACEHPKYSFTLEGTEVOP-IHEITD-FESPRQKPKSYRANGAIY 180
QY 213 ENG--SFYFAKPHLIE--MGYLQGGKWHHTKCELSHSDVIDIDWPIAEQRV 261
Db 181 INDIQSLFEKRFPPAPMRFFYL-----PTVERSIDIDSTLDQLAESLI 224
RESULT 12

PTMB CAMCO
ID PTMB CAMCO STANDARD; PRT; 235 AA.
AC Q45982; Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Posttranslational flagellin modification protein B.
DN PTMB.
OS Campylobacter coli.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
CX NCBI_TaxID=195;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VCL167;
RX MEDLINE=96423180; PubMed=8825781;
RA Guerry P., Doig P., Alm R.A., Burr D.H., Kinsella N., Trust T.J.;
RT "Identification and characterization of genes required for post-translational modification of Campylobacter coli VCL167 flagellin.";
RL Mol. Microbiol. 19:369-378 (1996).
CC -!- FUNCTION: REQUIRED FOR BIOSYNTHESIS OF LAH MODIFICATION IN THE POST-TRANSLATIONAL MODIFICATION OF CAMPYLOBACTER COLI FLAGELLIN.
CC -!- SIMILARITY: Belongs to the CMP-NeuNAc synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; U25992; AAB48074.1; --
CC PIR; S70685; S70685.
CC InterPro; IPR003329; Cytidylyl_trans.
CC Pfam; PF02348; CTP_transf_3; 1.
SQ SEQUENCE 235 AA; 26600 MW; A58421F922DA040B CRC64;
Query Match 7.5%; Score 171; DB 1; Length 235;
Best Local Similarity 26.0%; Pred. No. 2.5e-06;
Matches 63; Conservative 50; Mismatches 99; Indels 30; Gaps 10;
QY 49 ILARGSGKGIPLKNIKHLGVLIGWVLRALDAGAFOSVWVSTHDEIENKAKQF 108
Db 8 ICARGSGKGVKNKIRKINDLEMIAYSIQAKNGKLFKHIVISTDSEIATKALYGGV 67
QY 109 -HRSSEVSKDSSTSL-----DAIEFLNXXEXDIVGNIGTSCXCLHPTDLQKVAEMIR 162
Db 68 FFKREAHLASATAKIPVWRQALLRSEYFFKQCFDILLDASAPLRSSADIIKAFETFC 127
QY 163 BEGYDSXFSVV--RRHQFRMSEIQGVRETEPLNINPAKR---PREQDWDGELYENG 216
Db 128 QNQNDLITAVPARRNPYFNILIEVQDG--KVVKSGKNFTTRQSVKPCYDM-----NAS 179
QY 217 PFYAKRHLEMGYLOGGKWHHTKCELSHSDVIDIDWPIAEQRVLYGVFGKELKE 274
Db 180 IYIFKRDPELLQDCSVFGK-NTGLFIMDESTAFDVSSELDKIVE-----FLIKEKNLQ 232
QY 275 IK 276
Db 232 AK 233
RESULT 13
KDSB PASMU
ID KDSB PASMU STANDARD; PRT; 258 AA.
AC P57883;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-deoxy-manno-oculosonate cytidyltransferase (EC 2.7.7.38) (CMP-KDO synthetase) (CMP-2-Reto-3-deoxyoctulosonic acid synthetase) (CKS).
DN KDSB OR PWC859.

KDSB HAETN
ID KDSB HAETN STANDARD; PRT; 253 AA.
AC P44490;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-deoxy-manno-oculosonate cytidyltransferase (EC 2.7.7.38) (CMP-KDO
synthetase) (CMP-2-keto-3-deoxyoculosonic acid synthetase) (CKS).
GN KDSB OR H10058.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI TaxID=727;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd";
RL Science 269:496-512(1995).
CC -!- FUNCTION: Activates KDO (a required 8-carbon sugar) for
CC incorporation into bacterial lipopolysaccharide in Gram-negative
CC bacteria (By similarity).
CC -!- CATALYTIC ACTIVITY: CTP + 3-deoxy-D-manno-oculosonate =
CC diphosphate + CMP-3-deoxy-D-manno-oculosonate.
CC -!- PATHWAY: Lipopolysaccharide biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the kdsB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32691; AAC21736.1; -.
CC F1R; G64045; G64045.
CC T1GR; H10058; -.
CC HAVAP; MF 00057; -; 1.
CC InterPro; IPR003329; Cytidylyl_trans.
CC InterPro; IPR004528; KdsB.
CC Pfam; PF02348; CTP_transf_3; 1.
CC TIGRFAMs; TIGR00466; KdsB; 1.
CC Lipopolysaccharide biosynthesis; Transferase; Nucleotidyltransferase;
CC Complete proteome.
CC INIT MET 0 BY SIMILARITY.
CC SEQUENCE 253 AA; 28124 MW; 2F95E2B15CB56A2 CRC64;
Query Match 5.5%; Score 131; DB 1; Length 253;
Best Local Similarity 28.6%; Pred. No. 0.0041;
Matches 50; Conservative 32; Mismatches 69; Indels 24; Gaps 7;
QY 48 LILARGSGKPLKNIKHLGVLPLIGVLRALDSDGAFQSVWVSTDDHDEIENNAKQPCQAQ 107
DB 5 LIPARFASRLPGKPLADIKGKPMIQVTFKALQSGA-SRVIIADNENNAVAKSFGAE 63
QY 108 YHRASSEYKDSSTG-LDAIEFFLNXXNEXDVGNIQATSCXCLHPTDLQKVAEMIREEGY 166
DB 64 VCMTS--LYHNSGTIERLAEVVEKL-AIPDNEIIVWIGDEPLIPVIVRQVADNLAKENV 120
QY 167 DSXFSVTRHQRFRSEIOKGVREYVEPLNLPAPKPRQDWDGELYENGSGFYFAK 221

Db 121 NMASLAVKIH-----DAELENFNAVK--VLTKDGYV-----LYFSR 156

Search completed: May 6, 2004, 09:05:00
Job time : 13.4127 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Rur on: May 6, 2004, 08:59:19 ; Search time 42.4806 Seconds
(without alignments)
3223.466 Million cell updates/sec

Title: US-09-930-440B-4

Perfect score: 2266

Sequence: 1 MDSVEKGAATSVNPRGRPS.....REFAEHICILMEKVNNSCQK 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPREMBL.25.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_phase.*
- 11: sp_plant.*
- 12: sp_rodent.*
- 13: sp_virus.*
- 14: sp_vertebrate.*
- 15: sp_unclassified.*
- 16: sp_rvirus.*
- 17: sp_bacteriap.*
- 18: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Length	ID	Description
1	2198	97.0	434	Q8NFW8	Q8ntw8 homo sapien
2	2188	96.6	434	Q9NQZ0	Q9nqz0 homo sapien
3	2052	90.6	432	Q8K2G7	Q8k2g7 mus musculus
4	2048	90.4	432	Q88719	Q88719 mus musculus
5	2299	57.3	263	Q96AX5	Q96ax5 homo sapien
6	1107.5	48.9	432	Q90WG6	Q90wg6 oncofrynchu
7	952	42.0	200	Q8C330	Q8c330 mus musculus
8	813	35.9	166	Q99KK2	Q99kk2 mus musculus
9	552	24.4	330	Q8TJL6	Q8tjl6 methanosarc
10	402	17.7	424	Q9AK46	Q9ak46 streptomyce
11	376	16.6	432	Q82HY3	Q82hy3 streptomyce
12	288.5	12.7	234	Q8KDA1	Q8kda1 chlorobium
13	268	11.8	187	Q8IQV0	Q8iqv0 chroococcila
14	265	11.7	228	Q8F5R2	Q8f5r2 leptospira
15	250	11.0	229	Q89HL8	Q89hl8 bradyrhizob
16	231.5	10.2	179	Q87WU4	Q87w4 pseudomonas

17	230.5	10.2	163	16	Q67920	Q67920 aquifex aeo
18	228	10.1	233	16	Q8DE02	Q8de02 vibrio vuln
19	226.5	10.0	227	2	Q9R9M4	Q9r9m4 rhizobium m
20	225.5	10.0	169	2	Q9F2B5	Q9f2b5 thauera aro
21	223.5	9.9	174	2	Q8KLM5	Q8klw5 pseudomonas
22	223	9.8	209	16	Q8F5Q2	Q8f5q2 leptospira
23	223	9.8	421	2	Q93N01	Q93nd1 escherichia
24	217.5	9.6	164	16	Q8A712	Q8a712 bacteroides
25	217.5	9.6	188	16	Q7U9R7	Q7u9r7 synechococc
26	211.5	9.3	174	16	Q88P96	Q88p96 pseudomonas
27	209.5	9.2	179	16	Q8HV99	Q8hv99 pseudomonas
28	209	9.2	239	16	Q87T67	Q87t67 vibrio para
29	205.5	9.1	232	2	Q8BDX4	Q8bdx4 legionella
30	205	9.0	178	16	Q8JUS3	Q8jsu3 neisseria m
31	203.5	9.0	183	16	Q8EAE9	Q8eae9 shewarella
32	203.5	9.0	228	2	Q9R9S4	Q9r9s4 aeromonas p
33	203.5	9.0	413	2	Q9ALW4	Q9alw4 streptococc
34	202.5	8.9	229	2	Q8KHC9	Q8khc9 pseudomonas
35	202	8.9	226	16	Q7U960	Q7u960 synechococc
36	201.5	8.9	413	2	Q93T10	Q93ti0 streptococc
37	200	8.8	164	16	Q26090	Q26090 helicobacte
38	199.5	8.8	164	16	Q9ZJ39	Q9zj39 helicobacte
39	199	8.8	172	16	Q8KE66	Q8ke66 chlorobium
40	199	8.8	238	16	Q82UC4	Q82uc4 nitrosomona
41	197	8.7	168	16	Q8EGS1	Q8egs1 fusobacteri
42	194	8.6	178	16	Q9KL37	Q9kl37 neisseria m
43	193	8.5	223	16	Q9CP68	Q9cp68 pasteurella
44	192.5	8.5	165	16	Q83DI2	Q83di2 coxiella br
45	187	8.3	233	16	Q8RLF1	Q8rlf1 bradyrhizob

ALIGNMENTS

RESULT 1

ID	Q8NFW8	PRELIMINARY;	PRT;	434 AA.
AC	Q8NFW8;			
DT	01-OCT-2002 (TREMBLrel. 22, Created)			
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)			
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)			
DE	Cytidine monophospho-sialic acid synthase.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RA	Lawrence S.M., Huddleston K.A., Tomiya N., Nguyen N., Lee Y.C.,			
RA	Vann W.F., Coleman T.A., Betenbaugh M.J.;			
RT	"Cloning and Expression of Human Sialic Acid Pathway Genes to Generate			
RT	CMP-Sialic Acids in Insect Cells.";			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF397212; AAM90580.1; -			
DR	Genew; HGNC:18290; CMAS.			
DR	G2; GO:0039103; P:lipopolysaccharide biosynthesis; IEA.			
DR	InterPro; IPR003329; Cytidylyl_trans.			
DR	Pfam; PF02348; CTP_transf_3; 1.			
SQ	SEQUENCE 434 AA; 48449 MW; C8FAD3D8CDCF835B CRC64;			
Query Match 97.0%; Score 2198; DB 4; Length 434;				
Best Local Similarity 98.2%; Pred. No. 2.4e-173;				
Matches 426; Conservative 1; Mismatches 7; Indels 0; Gaps 0;				
QY	1	MDSVEKGAATSVNPRGRPSGRPPKLRNSRGQGRGVKPPHPLAALILARGSKGIPL	60	
Db	1	MDSVEKGAATSVNPRGRPSGRPPKLRNSRGQGRGVKPPHPLAALILARGSKGIPL	60	
QY	61	KNKHLAGVPLIGWVLRALDSCAFQSVWYSTDHDEIENVAKQFGAQRHRSRSEVSKSS	120	
Db	61	KNKHLAGVPLIGWVLRALDSCAFQSVWYSTDHDEIENVAKQFGAQRHRSRSEVSKSS	120	
QY	121	TSLDAIEFLNXXNCDIVGNIQATSCMLHTDLQKVAEMIREEGYDSXFSVVRHQFRW	180	


```

Db      12:  TSLDAIIFLYNXXEXDIVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180
QY      181  SEIQGVRETEPLNPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
Db      181  SEIQGVRETEPLNPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
QY      241  ELESVDIDVDIDPPIAQRVLRYGYFGKEKKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
Db      241  RAESVDIDVDIDPPIAQRVLRYGYFGKEKKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
QY      301  ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSLKLDCXEVSVSDKLAVDVDEWKE 360
Db      301  ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSLKLDCXEVSVSDKLAVDVDEWKE 360
QY      361  MGLCWKEVAYLGNVSDDECLKRVLSGAPADACSYAKAVGYICKNGRGGAIRFAEH 420
Db      361  MGLCWKEVAYLGNVSDDECLKRVLSGAPADACSYAKAVGYICKNGRGGAIRFAEH 420
QY      421  ICLMEKVNNSCQK 434
Db      421  ICLMEKVNNSCQK 434

RESULT 2
Q9NQ20 PRELIMINARY; PRT; 434 AA.
AC Q9NQ20;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE CMP-N-acetylneuraminic acid synthase (EC 2.7.7.43).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bouquin T., Mundy J.;
RT "Human mRNA for CMP-N-acetylneuraminic acid synthase.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Isoqai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Pulii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saico K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuto Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF271388; AAF76203.1; -.
DR EMBL; AK022927; BAB14311.1; -.
DR GO; GO:0006781; F:N-acylneuraminate cytidylyltransferase acti. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003329; Cytidylyl trans.
DR InterPro; IPR000794; Ketoacyl synth.
DR Pfam; PF02348; CTP transf 3; 1.
DR PROSITE; PS00606; S_KETOACYL SYNTHASE; 1.
KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 434 AA; 48379 MW; B303B6647EF81A3A CRC64;

Query Match 96.6%; Score 2188; DB 4; Length 434;
Best Local Similarity 97.0%; Pred. No. 1.6e-172;
Matches 421; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVSNPRGSRGPRPKLQNSRGQGGRGVEKPPHPLAAILARGSGKGIPL 60
Db 1 MDSVEKGAATSVSNPRGSRGPRPKLQNSRGQGGRGVEKPPHPLAAILARGSGKGIPL 60
QY 61 KNKHLACVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAKQFGAQVHRSSESVSKDSS 120

```

```

61 KNKHLACVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAKQFGAQVHRSSESVSKDSS 120
QY 121 TSLDAIIFLYNXXEXDIVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180
Db 121 TSLDAIIFLYNXXEXDIVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180
QY 181 SEIQGVRETEPLNPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
Db 181 SEIQGVRETEPLNPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
QY 241 ELESVDIDVDIDPPIAQRVLRYGYFGKEKKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
Db 241 RAESVDIDVDIDPPIAQRVLRYGYFGKEKKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
QY 301 ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSLKLDCXEVSVSDKLAVDVDEWKE 360
Db 301 ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSLKLDCXEVSVSDKLAVDVDEWKE 360
QY 361 MGLCWKEVAYLGNVSDDECLKRVLSGAPADACSYAKAVGYICKNGRGGAIRFAEH 420
Db 361 MGLCWKEVAYLGNVSDDECLKRVLSGAPADACSYAKAVGYICKNGRGGAIRFAEH 420
QY 421 ICLMEKVNNSCQK 434
Db 421 ICLMEKVNNSCQK 434

RESULT 3
Q8K2G7 PRELIMINARY; PRT; 432 AA.
AC Q8K2G7;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Cytidine monophospho-N-acetylneuraminic acid synthetase.
GN CMAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031500; AAH31500.1; -.
DR MGD; MGI:3337124; Cmas.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003329; Cytidylyl trans.
DR Pfam; PF02348; CTP transf 3; 1.
SQ SEQUENCE 432 AA; 48030 MW; 404CEB2302AB6A47 CRC64;

Query Match 90.6%; Score 2052; DB 11; Length 432;
Best Local Similarity 91.0%; Pred. No. 2.9e-161;
Matches 395; Conservative 12; Mismatches 25; Indels 2; Gaps 1;

QY 1 MDSVEKGAATSVSNPRGSRGPRPKLQNSRGQGGRGVEKPPHPLAAILARGSGKGIPL 60
Db 1 MDALEKGAATSGAPRGSRGPRPKLQNSR--GAGRGLEKPPHPLAALVLRARGSGKGIPL 58
QY 61 KNKHLACVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAKQFGAQVHRSSESVSKDSS 120
Db 59 KNIKELACVPLIGWVLRALDAGVFQSVVWSTDHDEIENVAKQFGAQVHRSSETSKJSS 118
QY 121 TSLDAIIFLYNXXEXDIVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180
Db 119 TSLDAIIFLYNXXEXDIVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 178
QY 181 SEIQGVRETEPLNPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
Db 179 SEIQGVRETEPLNPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 238
QY 241 ELESVDIDVDIDPPIAQRVLRYGYFGKEKKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300

```

```

Db      239 RAHNSVDIDVDIDWPIAQRVURFGYGEKELKEIKLVNIDGCTNGHHIYVSGDQKE- 298
QY      301 ISYDVKDAIGISLLKKSGLIEVLISERACSKQTLSSKLDCKMEVSVSDKLAIVDDEWRKE 360
Db      299 ISYDVKDAIGISLLKKSGLIEVLISERACSKQTLSSKLDCKTEVSVSDKLAIVDDEWRKE 358
QY      361 MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGGRGAIRFAEH 420
Db      359 MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGGRGAIRFAEH 418
QY      421 ICLLMKKNNSCK 434
Db      419 IFLLEKKNNSCK 432

RESULT 4
Q96AX5 PRELIMINARY; PRT; 432 AA.
ID      088719
AC      088719
DT      01-NOV-1998 (Tremblrel. 08, Created)
DT      01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT      01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE      CMP-N-acetylneuraminic acid synthetase (EC 2.7.7.43).
GN      CMAS.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98356124; PubMed=9689047;
RA      Mueser A.K., Eckhardt M., Potvin B., Muhlenhoff M., Stanley P.,
RA      Gerardy-Schahn R.;
RT      "Mammalian cytidine 5 (prime)-monophosphate N-acetylneuraminic acid
RT      synthetase: A nuclear protein with evolutionarily conserved structural
RT      motifs.";
RL      Proc. Natl. Acad. Sci. U.S.A. 95:9140-9145 (1998).
DR      EMBL; AF062215; C98915.1; -.
DR      MGD; MGI:1337124; Cmas.
DR      GO; GO:0008781; P:N-acetylneuraminat cytidyltransferase acti. .; IEA.
DR      GO; GO:0016740; P:transferase activity; IEA.
DR      GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR      InterPro; IPR003329; Cytidylyl trans.
DR      Pfam; PF02348; CTP_transf_3; 1.
KW      Nucleotidyltransferase; Transferase.
SQ      SEQUENCE 432 AA; 48058 MW; BF7295535E7F6CE5 CRC64;

Query Match          90.4%; Score 2048; DB 11; Length 432;
Best Local Similarity 90.8%; Pred. No. 6.3e-161;
Matches 394; Conservative 12; Mismatches 26; Indels 2; Gaps 1;

QY      1 MDSVEKGAATSVNPRGRPSRPPKLRNSGGQGRGVEKPPHLLAALILARGSGKGIPL 60
Db      1 MDALEKGAATVSGPAPRGRPSRPPKLRNSR--GAGRGLEKPPHLLAALVLRGSGKGIPL 58
QY      61 KNIKHLAGVPLIGWLRAALDSGAFQSVWVSTHDEIENKAVQFGAQRHRSSEVSKDSS 120
Db      59 KNIKHLAGVPLIGWLRAALDSGAFQSVWVSTHDEIENKAVQFGAQRHRSSEVSKDSS 118
QY      121 TSLDAIIEFLNKNEXDIVGNIQATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHQRFW 180
Db      119 TSLDAIVFELNKNEXDIVGNIQATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHQRFW 178
QY      181 SEIQGVREVTEPLNLNPAKRPRQDWGELYENGSGFYFAKRLHIEMGYLQGGKWHHTKC 240
Db      179 SEIQGVREVTEPLNLNPAKRPRQDWGELYENGSGFYFAKRLHIEMGYLQGGKWHHTKC 238
QY      241 ELEHSVDIDVDIDWPIAQRVURFGYGEKELKEIKLVNIDGCTNGHHIYVSGDQKEI 300
Db      239 RAHNSVDIDVDIDWPIAQRVURFGYGEKELKEIKLVNIDGCTNGHHIYVSGDQKEI 298
QY      301 ISYDVKDAIGISLLKKSGLIEVLISERACSKQTLSSKLDCKMEVSVSDKLAIVDDEWRKE 360

```

```

Db      299 ISYDVKDAIGISLLKKSGLIEVLISERACSKQTLSSKLDCKTEVSVSDKLAIVDDEWRKE 358
QY      361 MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGGRGAIRFAEH 420
Db      359 MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGGRGAIRFAEH 418
QY      421 ICLLMKKNNSCK 434
Db      419 IFLLEKKNNSCK 432

RESULT 5
Q96AX5 PRELIMINARY; PRT; 263 AA.
ID      Q96AX5
AC      Q96AX5;
DT      01-DEC-2001 (Tremblrel. 19, Created)
DT      01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT      01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE      Similar to CMP-N-acetylneuraminic acid synthase.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      TISSUE=Placenta;
RA      Strausberg R.;
RL      Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC016609; AAH16609.1; -.
DR      GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR      InterPro; IPR003329; Cytidylyl trans.
DR      Pfam; PF02348; CTP_transf_3; 1.
SQ      SEQUENCE 263 AA; 29561 MW; 5A51E1F0C98ABIDB CRC64;

Query Match          57.3%; Score 1299; DB 4; Length 263;
Best Local Similarity 95.4%; Pred. No. 3e-99;
Matches 251; Conservative 1; Mismatches 11; Indels 0; Gaps 3;

QY      1 MDSVEKGAATSVNPRGRPSRPPKLRNSGGQGRGVEKPPHLLAALILARGSGKGIPL 60
Db      1 MDSVEKGAATSVNPRGRPSRPPKLRNSGGQGRGVEKPPHLLAALILARGSGKGIPL 60
QY      61 KNIKHLAGVPLIGWLRAALDSGAFQSVWVSTHDEIENKAVQFGAQRHRSSEVSKDSS 120
Db      61 KNIKHLAGVPLIGWLRAALDSGAFQSVWVSTHDEIENKAVQFGAQRHRSSEVSKDSS 120
QY      121 TSLDAIIEFLNKNEXDIVGNIQATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHQRFW 180
Db      121 TSLDAIIEFLNKNEXDIVGNIQATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHQRFW 180
QY      181 SEIQGVREVTEPLNLNPAKRPRQDWGELYENGSGFYFAKRLHIEMGYLQGGKWHHTKC 240
Db      181 SEIQGVREVTEPLNLNPAKRPRQDWGELYENGSGFYFAKRLHIEMGYLQGGKWHHTKC 240
QY      241 ELEHSVDIDVDIDWPIAQRVURFGYGEKELKEIKLVNIDGCTNGHHIYVSGDQKEI 300
Db      241 RAHNSVDIDVDIDWPIAQRVURFGYGEKELKEIKLVNIDGCTNGHHIYVSGDQKEI 298

RESULT 6
Q96AX5 PRELIMINARY; PRT; 432 AA.
ID      Q96AX5
AC      Q96AX5;
DT      01-DEC-2001 (Tremblrel. 19, Created)
DT      01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT      01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE      CMP-sialic acid synthetase.
OS      Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC      Prclacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX      NCBI_TaxID=8022;
RN      [1]

```

```

SEQUENCE FROM N.A.
RP TISSUE=Testis;
RC MEDLINE=21372051; PubMed=11479279;
RA Nakata D., Munster A.K., Gerardy-Schahn R., Aoki N., Matsuda T.,
RA Kitajima K.;
RT "Molecular cloning of a unique CMP-sialic acid synthetase that
RT effectively utilizes both deaminoneuraminic acid (XDN) and N-
RT acetylneuraminic acid (Neu5Ac) as substrates.";
RL Glycobiology 11:685-692(2001).
DR EMBL: AB027414; BAB47150.1; -.
DR CC: GO:0005123; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro: IPR003329; Cytidylyl trans.
DR Pfam: PF02348; CTP_transf_3; 1.
SQ SEQUENCE 432 AA; 48242 MW; 51947AC257B23382 CRC64;

Query Match 48.9%; Score 1107.5; DB 13; Length 432;
Best Local Similarity 54.3%; Pred. No. 4.2e-83;
Matches 227; Conservative 59; Mismatches 119; Indels 13; Gaps 5;

QY 21 RGRPEKLGNSGGQSGRGVEXKPPHIALALILARGGSKGIPLNKIKELAGVPLTGWVIRAL 80
DB 16 RDRKAKVIDS--GSKR-----HTAAILARGGSKG-PLNKKIKVLAVGLLIGVLRVAAV 67

QY 81 DSGAFQSVWSTDDHDEIENKAKQFQAQVRRSRSEVSKSSSLDAIIEFLNYXNEDIVG 140
DB 68 DSKQFDSVWSTDDHDEIEKAKTGAQVRRSRSEVSKSSSLDTIQEFARLNPEVDVIC 127

QY 141 NIQATXCLHPTDLOKVAEMIREEGYDSXFSVVRHQRWSELOKSVRETEPLNLNPAK 200
DB 128 HIQATSPCLHFPHLKEALEMTIKQFTSVFVRRHHFWQVKKGGVATQPLNLDPCN 187

QY 201 RPRQDWDGELYNGSFYFAKHLHLEMGYLOGGKWHHTTCKEHSVDIDWDIDWPAQR 260
DB 188 RPRQDWDGELCENGSEFYITRATIERG-LQGGKWAYVEMLEPEYSVDIDWDIDWPAQR 246

QY 261 VLRGYFGKEKLEKLIVCNIDGCLTNGEIVYVSDQKEIISYDVKDAIGISLLKKSIE 320
DB 247 VLRGYFGKLEPEVRLNLCNVSGCLTDGRVLISVSGEEMVSNTRDTWGRMLQREGVE 306

QY 321 VRLT--SRACKSKQTLSSL--KLCKMEVSVSDKLAVDWEKMGKQKWEVAYLGNVS 376
DB 307 VILSSSEDLTKALADNLSTQRTGCEVRLQSKDIQGEVIAMDDKDLWKEVAYVGNDA 366

QY 377 DEECLKRVKLGSGAPADACSYAQKAVGYICKNGRGGAIRFAEHICLLMEKVNNSCQK 434
DB 367 DVDCLNLAGLSAVPRDAPWAINAAKYSCHSAAGLGAVERFSEHILLKKAQKQMEQ 424

RESULT 7
Q8C330 PRELIMINARY; PRT; 200 AA.
AC Q8C330;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytidine monophospho-N-acetylneuraminic acid synthetase.
GN CMAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL: AK087150; BAC39813.1; -.
DR MGD; MGI:1337124; Cmas.
DR GO: GO:0009103; P:lipopolysaccharide biosynthesis; IEA.

DR InterPro: IPR003329; Cytidylyl trans.
DR Pfam: PF02348; CTP_transf_3; 1.
SQ SEQUENCE 200 AA; 22276 MW; 0D3C036CD8ECAF31 CRC64;

Query Match 42.0%; Score 952; DB 11; Length 200;
Best Local Similarity 94.8%; Pred. No. 1e-70;
Matches 182; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 243 EHSVVDIDWDIDWPIAEQVLRVYGYFGKEKLEKLIVCNIDGCLTNGHIVYVSDQKEIIS 302
DB 9 EHSVVDIDWDIDWPIAEQVLRVYGYFGKEKLEKLIVCNIDGCLTNGHIVYVSDQKEIIS 68

QY 303 YDVKDAIGISLLKKSIEVRLTISERACSKQTLSALKLCKTEVSVSDKLAVDWEKMG 362
DB 69 YDVKDAIGISLLKKSIEVRLTISERACSKQTLSALKLCKTEVSVSDKLAVDWEKMG 128

QY 363 LCKKEVAYLGNVSDEECLKRVKLGSGAPADACSYAQKAVGYICKNGRGGAIRFAEHIC 422
DB 129 LCKKEVAYLGNVSDEECLKRVKLGSAVPADACGAKAVGYICKSGRGGAIRFAEHIF 198

QY 423 LLMEKVNNSCQK 434
DB 189 LLIEKVNNSCQK 200

RESULT 8
Q99KK2 PRELIMINARY; PRT; 166 AA.
AC Q99KK2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to cytidine monophospho-N-acetylneuraminic acid synthetase
DE (Fragment).
GN CMAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004606; AA04606.1; -.
DR MGD; MGI:1337124; Cmas.
DR InterPro: IPR008230; Sugar_Phase.
DR PIRSF; PIRSF006118; Sugar_Phase; 1.
FT NON_TER 1
SQ SEQUENCE 166 AA; 18201 MW; B70B2B32F38DA3D8 CRC64;

Query Match 35.9%; Score 813; DB 11; Length 166;
Best Local Similarity 94.6%; Pred. No. 2.6e-59;
Matches 157; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 269 KEKLEKLIVCNIDGCLTNGHIVYVSDQKEIISYDVKDAIGISLLKKSIEVRLISRA 328
DB 1 KEKLEKLIVCNIDGCLTNGHIVYVSDQKEIISYDVKDAIGISLLKKSIEVRLISRA 60

QY 329 CSKQTLASLKLDCKMEVSVSDKLAVDWEKMGKQKWEVAYLGNVSDEECLKRVKLGSG 388
DB 61 CSKQTLASLKLDCKTEVSVSDKLAVDWEKMGKQKWEVAYLGNVSDEECLKRVKLGSA 120

QY 389 APADACSYAQKAVGYICKNGRGGAIRFAEHICLLMEKVNNSCQK 434
DB 121 VPADACSGAKAVGYICKSGRGGAIRFAEHIFLLIEKVNNSCQK 166

RESULT 9
Q8TJL6 PRELIMINARY; PRT; 390 AA.
AC Q8TJL6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

```


Db 120 TSPFLWEDVGVAGVVEDGCTAVTAPVPHGFWRDGADEPGAGGCHVWHDKVRP 179
 Qy 203 RRQMDGELYENGSGFY-----FAK-RHLIEMGYLOGGKWHHTKCELESHSDVIDDIDWP 255
 Db 180 RRQRPQDLLETGAAYAAVAPGFRKHRRF-----FGRTDLVRTDPARVLEIDDPHD-- 231
 Qy 256 IAEORVLYRYGFGKEK-----LKEIKLVNIDGCLTNGHHIYSGDQKEIISYDVKDAL 309
 Db 232 LAARALAA-PHFDTARFPAALPTAADIIDAVVLDFDGTQDDRVLDITDGRFEVSVHRGDGL 290
 Qy 310 GISLLKSGIEVRLISERACSKQCLSSIKLDCKEVSVSDKLAIVDWEKEMGLCWKEVA 369
 Db 291 GIALRRSGLTMLILSTEVPVVAARAKKLPLVHGHIDRKLALQKWCCEQGIAPERV 350
 Qy 370 YLGNVSEDECKLVSGAPADACSVAKVAGVICKNGGRGAIRFPA 418
 Db 351 YVGNVDNDLPCFALVGVFVAVASAHDAVRGAARAVTTVGGGAVREIA 399

RESULT 11

Q82HY3 ID Q82HY3 PRELIMINARY; PRT; 432 AA.
 AC Q82HY3;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Putative transferase.
 GN SAV3374.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomyces; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=2:477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M., Murata S.,
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692362;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Murata S.,
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis."
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AP005034; BAC71086.1; -
 DR GO; GO:0006740; P:transferase activity; IEA.
 DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
 DR InterPro; IPR003329; Cytidyl transferase.
 DR InterPro; IPR008230; Sugar ptase.
 DR Pfam; PF02348; CTP_transf_3; 1.
 DR PIRSF; PIRSF06118; Sugar ptase; 1.
 KW Transferase; Complete prctome.
 SQ SEQUENCE 432 AA; 45461 MW; 053CC8C8C17D94DB CRC64;

Query Match 16.6%; Score 376; DB 16; Length 432;
 Best Local Similarity 27.4%; Pred. No. 1:5e-22;
 Matches 118; Conservative 75; Mismatches 188; Indels 50; Gaps 9;
 Qy 30 NSRGGQGRVGPPLAAILIAGSGSGTFLKNKHLGVLGWLRAALDSGAFQSVW 89
 Db 3 HSEAGCAASVR---VLAVIPAGSGSGVPAKNLLPVGVPLVARAVRECRATLVTV 59
 Qy 90 VSTDHDE:ENVAKQFGAOWHRSSEVSKDSSTSLDAIEFLN-----YXNEXDIVGN:QA 144

Db 50 VSTDHDAIAAARAGAEVVLPAIAAGTATSEAAVLHMDTHEALEGAAVVDVLLVQC 119
 Qy 145 TSXCLHPTDLOKVAEMIREGYDSKFSVVRHQRWSEI-----QKGVREVTSEL 194
 Db 120 TSPFVREVDVGVVHAIKVGADTALTAVPHFGVWRDADDPALGALGAERAAVEGDTIL 179
 Qy 195 -----NLNPAKRRPQDWDGELVNGSFYPAKHLIEMGYLOG-----GKWH 237
 Db 180 VTGTATSGGGINHDKSFRPRQRDPQDLLETGAVY-----GMDATGFEARHFRTEL 235
 Qy 238 TKCELESHSDVIDDIDWPIAEQRLV-----RYGYFQKEKIKLKVNCNIDGCLTNGH 290
 Db 236 VRTDPARVLEIDDPHD--LAARALAPLFDANRPG--ALPTAEDIDAVVLDFDGTQDDR 291
 Qy 291 IYVSGDQKEIISYDVKDAIGISILKSGIEVRLISERACSKQCLSSIKLDCKEVSVSDK 350
 Db 292 VLIDSDGREFVSVHRGDGLGIAALKKSLTLLIISTEONPVVAARAKKLPLVHGHIDR 351
 Qy 351 LAVVDEWRKEMGLCWKEVAYELGNEVSEDECKLVKSGAPADACSVAKVAGVICKNGG 410
 Db 352 DLALKQWCCEQGIAPERVLYVGNVDNDLPCFALVGVFVAVASAHDAVRGAARAVTTVP 411
 Qy 411 RGAIRFPAHI 421
 Db 412 DGAIRETASWI 422

RESULT 12

Q8KDAl ID Q8KDAl PRELIMINARY; PRT; 234 AA.
 AC Q8KDAl;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Acylneuraminate cytidyl transferase.
 GN NEUA OR CT1153.
 OS Chlorobium tepidum.
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
 OC Chlorobium.
 OX NCBI_TaxID=1097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TLS / ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
 RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
 RA Holt I., Umayam L.A., Mason T., Brenner M., Stea T.P., Parksey D.,
 RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
 RA Vanatavan J., Khouiri H., White O., Gruber T.M., Ketchum K.A.,
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.,
 RT "The complete genome sequence of Chlorobium tepidum TLS, a
 RT photosynthetic, anaerobic, green-sulfur bacterium."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
 DR EMBL; AB012876; AAM72386.1; -
 DR TIGR; CT1153; -
 DR GO; GO:0016779; F:nucleotidyltransferase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0039103; P:lipopolysaccharide biosynthesis; IEA.
 DR InterPro; IPR003329; Cytidyl transferase.
 DR Pfam; PF02348; CTP_transf_3; 1.
 KW Transferase; Nucleotidyltransferase; Complete proteome.
 SQ SEQUENCE 234 AA; 26382 MW; 38B7179E43D9C206 CRC64;

Query Match 12.7%; Score 288.5; DB 16; Length 234;
 Best Local Similarity 35.1%; Pred. No. 1:1e-15;
 Matches 79; Conservative 27; Mismatches 86; Indels 33; Gaps 5;
 Qy 46 AAILIARGSGKIPKIKHLGVLGWLRAALDSGAFQSVVYSTDHDE:ENVAKQFG 105
 Db 4 AAILIPARGSGKIKKNIHPHIGPLPILAVSLQALCAEHVDQVFTTDDAIAQVAREFG 63
 Qy 106 AQVHRSSEVSKDSSTSLDAIEFLN-----YXNEXDIVGN:QA 160

```

Db 64 AVEIDRPER--SGDKATSESAHLEAIVTAERYGABPETWFIQATSPLRKFGDIDRAIEL 123
Qy 161 IREEGYDSEKSWRRHQFR-----WSEIQKGWRETEPLNMLNPAKPRRQDWDGELY 212
Db 124 FRLEGANSLSVTRADDLTIWEQRGGDNVS-----NFYRNGRQDRPSOFI 172
Qy 213 ENGSPY-PAKHLEIMGYLGKK-----WHTTKCLEHSVDI 248
Db 173 ENGSIYLF-PSVLRREFGNRCQKLSVLMFWQIWEITIEVDL 217

RESULT 13
Q81QV0 PRELIMINARY; PRT; 187 AA.
AC Q81QV0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CG32220-PA.
GN CG32220.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Slazell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Cabot G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berner J.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.V., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler K., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Durkin B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann M.,
RA Rosier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobbart T.J., Moberg M.P., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reizert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2];
RP SEQUENCE FROM N.A.
RX Celniker S.E., Adams M.D., Krommiller B., War K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

```

```

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Howland T.J., Hostin D., Howland T.J.,
RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Kunco J.,
RA Paclab J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3];
RP SEQUENCE FROM N.A.
RX Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak P., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4];
RP SEQUENCE FROM N.A.
RX Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5];
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003515; AAN11643.1; -.
DR FlyBase; FBgn0052220; CG32220.
SQ SEQUENCE 187 AA; 21754 MW; 0B65CDE59DEC893E CRC64;

Query Match 11.8%; Score 268; DB 5; Length 187;
Best Local Similarity 34.1%; Pred. No. 3.8e-14;
Matches 61; Conservative 33; Mismatches 77; Indels 8; Gaps 3;

Qy 81 DSGAFQSVWVSTHDELENNAKQGAOVHRSSVSKDSTSLDAITFLNXXNEXDIWG 140
Db 5 NSTCFRIHWSTDDKRAIEAQAQKYGAIHHRPEKARDTSLFA--SEFLDVHRSIHDF 64
Qy 141 NIQATSCXCLHPTDLQKVAEMIRE-EGYDSXPSVVRHQFRWSEIQKGWRETEPLNMLNPA 199
Db 65 LFQCTSVFLKTKYIQ---EAVRKPESHDCVFAAKRSHYLRW---KVVYDGEIMPAEPDLS 117
Qy 200 KRPRQDQDGLVNGSPFYFAKRLHEIMGYLGKKWHTTKCLEHSVDIDWDPIAE 258
Db 118 ARPRQDQDGLVNGSPFYFAKRLHEIMGYLGKKWHTTKCLEHSVDIDWDPIAE 176

RESULT 14
Q8F5R2 PRELIMINARY; PRT; 228 AA.
AC Q8F5R2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NEU1 OR LA1605.
GN Leptospira interrogans.
OS Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1];
RP SEQUENCE FROM N.A.
RX STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB011338; AAN48804.1; -.
DR GO; GO:0008781; F:N-acylneuraminate cytidyltransferase acti. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003323; Cytidylyl_trans.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:59:54 ; Search time 18.0701 Seconds
(without alignments)
1239.929 Million cell updates/sec

Title: US-09-930-440B-4

Perfect score: 2266

Sequence: 1 MDSVEKGAATSVSNPRGRPS.....REFAEHICLLMEKVNNSCQK 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PTCUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2258	99.6	434	US-09-516-143A-2	Sequence 2, Appli
2	225.5	10.0	169	US-09-516-914-7	Sequence 7, Appli
3	216	9.5	194	US-09-543-681A-5902	Sequence 5902, Ap
4	215.5	9.5	255	US-09-489-039A-13663	Sequence 13663, A
5	213.5	9.4	377	US-09-252-991A-24747	Sequence 24747, A
6	197.5	8.7	177	US-09-540-236-3597	Sequence 3597, Ap
7	169.5	7.5	221	US-09-495-406-23	Sequence 23, Appl
8	159	7.0	180	US-09-328-352-6372	Sequence 6372, Ap
9	128	5.6	467	US-08-867-611-28	Sequence 28, Appl
10	128	5.6	467	US-09-690-359-28	Sequence 28, Appl
11	128	5.6	467	PCT-US92-06965A-33	Sequence 33, Appl
12	128	5.6	541	US-08-867-611-8	Sequence 8, Appli
13	128	5.6	541	US-09-690-359-8	Sequence 8, Appli
14	128	5.6	541	PCT-US92-06965A-13	Sequence 13, Appl
15	128	5.6	594	US-08-867-611-48	Sequence 48, Appl
16	128	5.6	594	US-09-690-359-48	Sequence 48, Appl
17	128	5.6	599	US-08-867-611-18	Sequence 18, Appl
18	128	5.6	599	US-09-690-359-18	Sequence 18, Appl
19	128	5.6	599	PCT-US92-06965A-23	Sequence 23, Appl
20	122	5.4	496	US-08-867-611-10	Sequence 10, Appl
21	122	5.4	496	US-09-690-359-10	Sequence 10, Appl
22	122	5.4	496	PCT-US92-06965A-15	Sequence 15, Appl
23	121	5.3	387	US-08-867-611-12	Sequence 12, Appl
24	121	5.3	387	US-09-690-359-12	Sequence 12, Appl
25	121	5.3	387	PCT-US92-06965A-17	Sequence 17, Appl
26	120	5.3	460	US-08-912-129A-60	Sequence 60, Appl
27	120	5.3	466	US-08-912-129A-55	Sequence 55, Appl

28	120	5.3	490	2	US-08-912-129A-50	Sequence 50, Appl
29	120	5.3	618	2	US-08-912-129A-54	Sequence 54, Appl
30	119.5	5.3	342	3	US-08-867-611-37	Sequence 37, Appl
31	119.5	5.3	342	4	US-09-690-359-37	Sequence 37, Appl
32	119.5	5.3	344	3	US-08-867-611-38	Sequence 38, Appl
33	119.5	5.3	344	4	US-09-690-359-38	Sequence 38, Appl
34	119.5	5.3	352	3	US-08-867-611-39	Sequence 39, Appl
35	119.5	5.3	352	4	US-09-690-359-39	Sequence 39, Appl
36	119.5	5.3	357	3	US-08-867-611-40	Sequence 40, Appl
37	119.5	5.3	357	4	US-09-690-359-40	Sequence 40, Appl
38	119.5	5.3	362	3	US-08-867-611-41	Sequence 41, Appl
39	119.5	5.3	362	4	US-09-690-359-41	Sequence 41, Appl
40	119.5	5.3	363	3	US-08-867-611-43	Sequence 43, Appl
41	119.5	5.3	363	4	US-09-690-359-43	Sequence 43, Appl
42	119.5	5.3	364	3	US-08-867-611-44	Sequence 44, Appl
43	119.5	5.3	364	4	US-09-690-359-44	Sequence 44, Appl
44	119.5	5.3	365	3	US-08-867-611-42	Sequence 42, Appl
45	119.5	5.3	365	4	US-09-690-359-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1
US-09-516-143A-2
; Sequence 2, Application US/09516143A
; Patent No. 6333182
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OR INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: P2505PCT
; CURRENT APPLICATION NUMBER: US/09/516.143A
; CURRENT FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: 133
; OTHER INFORMATION: Xaa equals Tyr or His
; NAME/KEY: SITE
; LOCATION: 136
; OTHER INFORMATION: Xaa equals Gly or Val
; NAME/KEY: SITE
; LOCATION: 147
; OTHER INFORMATION: Xaa equals Ser or Pro
; NAME/KEY: SITE
; LOCATION: 169
; OTHER INFORMATION: Xaa equals Gly or Val
US-09-516-143A-2

Query Match 99.6%; Score 2258; DB 4; Length 434;

Best Local Similarity 100.0%; Pred. No. 1.6e-223;

Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVSNPRGRPPKLRNSRGQGRGVEKPPHIALIILARGSGKGIPL 60

Db 1 MDSVEKGAATSVSNPRGRPPKLRNSRGQGRGVEKPPHIALIILARGSGKGIPL 60

QY 61 KNIKHLAGVPLIGWLRAALDSGAFQSVWYSTDHDTEINVAKQFGAOVHRRSSSVKDS 120

Db 61 KNIKHLAGVPLIGWLRAALDSGAFQSVWYSTDHDTEINVAKQFGAOVHRRSSSVKDS 120

QY 121 TSLDAIEFLNYKXNDIVGNIQATSKLHPTDLQKVAEMIREEGYDSXFSVVRHOPRW 180

Db 121 TSLDAIEFLNYKXNDIVGNIQATSKLHPTDLQKVAEMIREEGYDSXFSVVRHOPRW 180

QY 181 SETQKGVREVTPELNLPNAKPRQWDGELYNGSEFFAKRHLIEMGYLQGGKWHITTC 240

Db 181 SEIQGVREVEPLNPAKRFRQDWDGELYENGFFAKSHLIEMGYLQSGKWHITKC 240
QY 241 ELESVDIDVIDIDPIAQRVRYGYFGKBLKEIKLLVNCNIDGCLTNGHIYVSGDQKEI 300
Db 242 ELESVDIDVIDIDPIAQRVRYGYFGKBLKEIKLLVNCNIDGCLTNGHIYVSGDQKEI 300
QY 301 ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLAVDWPK 360
Db 301 ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLAVDWPK 360
QY 361 MGLCWKEVAYLNGEVSDEECLKRVLSGAPADACSYAKAVGYICKNGRGGAIRFAFH 420
Db 361 MGLCWKEVAYLNGEVSDEECLKRVLSGAPADACSYAKAVGYICKNGRGGAIRFAFH 420
QY 421 ICLLMKVNNSCK 434
Db 421 ICLLMKVNNSCK 434

RESULT 2

US-09-516-914-7
; Sequence 7, Application US/09516914
; Patent No. 6333401
; GENERAL INFORMATION:
; APPLICANT: Breinig, Sabine
; APPLICANT: Fuchs, Georg
; TITLE OF INVENTION: Phenol-induced Proteins of Thauera aromatica
; FILE REFERENCE: EC1006 US NA
; CURRENT APPLICATION NUMBER: US/09/516,914
; EARLIER FILING DATE: 2000-03-01
; EARLIER APPLICATION NUMBER: 60/122,952
; EARLIER FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Thauera aromatica
US-09-516-914-7

Query Match 10.0%; Score 225.5; DB 4; Length 169;
Best Local Similarity 31.2%; Pred. No. 4.8e-15;
Matches 48; Conservative 39; Mismatches 64; Indels 3; Gaps 2;
QY 270 EKLKEIKLVNCIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKSGIEVRLISER-- 327
Db 2 EQAKNIKLVLDVGVMTDGRIVINDEGIESRNFIDKGMGVIVLQCGVEVAIIITSKRS 61
QY 328 ACSKQTLSSKLDCKMEVSVSDKLAVDWPKMEGLCWKEVAYLNGEVSDEECLKRVGLS 387
Db 62 GAVRHRAEELKIK-RFHEGIIKKTEPYAQMLEEMNISDAEVCYVGDDLDVLSMKRVGLA 120
QY 388 GAPADACSYAKAVGYICKNGRGGAIRFAFH 421
Db 121 VAVGDAVDVKEAAYVTTARGGEGAVREVAELI 154

RESULT 3

US-09-543-681A-5902
; Sequence 5902, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5902

; LENGTH: 194
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5902
Query Match 9.5%; Score 216; DB 4; Length 194;
Best Local Similarity 29.8%; Pred. No. 5.8e-14;
Matches 53; Conservative 42; Mismatches 67; Indels 16; Gaps 3;
QY 264 YGYFGK---EKLKEIKLVNCIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKSGIE 320
Db 18 YGAVSKQIMQKAEKQVLLICVDGVNSDGLIYMGNGBELKAEFVRDGYGIRCLLTSGIE 77
QY 321 VRLISERACSKQTLSSKLDCKMEVSVS-----DKLAVDWPKMEGLCWKEVAYLNCN 373
Db 78 VAIITGRQ-----SKLEDRAKTIGITYLYGQENKLLAYQQLDITLNKPEQATAYIGD 131
QY 374 EVSDEECLKRVLSGAPADACSYAKAVGYICKNGRGGAIRFAFHICLLMEKVNNS 431
Db 132 DLIDLPMVEKVGSLVAVADAHELLTPRADYVTHILGGRGAVRELCDLLLAQGLBEA 189

RESULT 4

US-09-489-039A-13663
; Sequence 13663, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13663
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13663

Query Match 9.5%; Score 215.5; DB 4; Length 255;
Best Local Similarity 32.3%; Pred. No. 1e-13;
Matches 54; Conservative 36; Mismatches 68; Indels 9; Gaps 2;
QY 270 EKLKEIKLVNCIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKSGIEVRLISERAC 329
Db 88 DRAAKIRILLIDVGVLSGDLIYMGNGBELKAEFVRDGYGIRCLTSGIEVAIITGRKA 147
QY 330 SK-----QTLSSKLDCKMEVSVSDKLAVDWPKMEGLCWKEVAYLNGEVSDEECLKRV 384
Db 148 KLVREDRCOTLGITHI-----YQGSQKLLAFRDLTDKLVHREVEVAYIGDLDLDPVMAEV 203
QY 385 GLSGAPADACSYAKAVGYICKNGRGGAIRFAFHICLLMEKVNNS 431
Db 204 GLSVAVADAHPLLPRTVTVTRINGRGCAVREVCDDLLELLAQGLDEA 250

RESULT 5

US-09-252-991A-24747
; Sequence 24747, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24747
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24747

Query Match 9.4%; Score 213.5; DB 4; Length 377;
Best Local Similarity 28.0%; Pred. No. 3.1e-13;
Matches 73; Conservative 42; Mismatches 115; Indels 31; Gaps 7;
QY 175 RHQFRSEIYOK-GVREVTPLNLP-----KPRQWDGELYNGSYFAKRLIEM 227
Db 119 RHFRWRPZASRPGSHRRPFGHRRPQDRPRQPRNRNRGSPEDHGR-----QQ 169
QY 228 GYLQGG--KWHHTTKCELSVDI---DVIDWPIAEQVLRVYGFCKEKKIKLLVCN 281
Db 170 DRTGGRRGRPSGRPEHARPVARSQVMSDTHSAELPVR-----ARRIRLAIFD 221
QY 282 IDGCLTNGHIYVSGDQKEIISYDVKDAIGISLKKSGIEVRLISERACSKQTLSSLLKDC 341
Db 222 VDGVLTDGKLYPLVDSGEKFTFTLDGHIKMLIASGVRTAITITGRTPVVERARNLGI 281
QY 342 K-MEVSVDKLAIVDWRKEMGLCWKEVAYLNEVSDDECLKRVGLSGAPADACSAQKA 400
Db 282 QHLYQSRDKLAIVDELGLGIGYQVAYLGGDPLPVIIRVGLGMAVASADPEVROH 341
QY 401 VGVICKNGGGRGAIRFAEHI 421
Db 342 AHGVTAAARGEGAREFCELI 362

RESULT 6
US-09-540-236-3597
; Sequence 3597, Application US/09542236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709,2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3597
; LENGTH: 177
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-3597

Query Match 8.7%; Score 197.5; DB 4; Length 177;
Best Local Similarity 29.4%; Pred. No. 4e-12;
Matches 50; Conservative 38; Mismatches 63; Indels 19; Gaps 4;
QY 270 EKLKEIKLVNCDGCTNGHIYVSGDQKEIISYDVKDAIGISLKKSGIEVRLISERAC 329
Db 10 QKAKHKLPMADVGDGILSDGIIYNSGEIETKAFYVQDGLQALKQSGIILAITGR-- 67
QY 330 SKQTLSSLLKLDCK-MEVSYS-----DKLAVDWEKREMGKVCWEVAYLGNVEVSECKL 381
Db 68 -----SSAMVDRAXELGISHIIOGQNDKLTALVGLAKLGIHLSHCATIGDLPDKAV 122
QY 382 KRVGLSGAPADACSAQKAQVYCKNGGGRGAIRFAEHIICLLMEKVNNS 431
Db 123 REVGFGLSPNGCEQTRAVSDYITTTGNGAVRE----VCELLKAQNN -68

RESULT 7
US-09-495-406-23
; Sequence 23, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:

; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/495,406
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: CMP-sialic acid synthetase from C. jejuni OHA384
; OTHER INFORMATION: (ORF 10a of LOS biosynthesis locus)
US-09-495-406-23

Query Match 7.5%; Score 169.5; DB 4; Length 221;
Best Local Similarity 28.3%; Pred. No. 4.3e-09;
Matches 63; Conservative 35; Mismatches 106; Indels 19; Gaps 5;
QY 47 ALILARGSGKGIPLKNKHLGAVPLIGWILRAALDSGAFOSVWVSTDHDIENVAQFGA 106
Db 4 AILPARGSGKGIKNKLVLLNNKPLIYTTTAAALNTKSIKVVVSSDSDELNVAKQNV 63
QY 107 QVRRSRSEVSKOSTSLDALLEINLYNKEADIVNQCATSXCLHPTDLOKVAEMIREGY 166
Db 64 DILKRPISLAQDNTTSDKVLHALKFYKCYEVVFLQPTSPLTRTNHIDEAFNLYKNSA 123
QY 167 DSXFSVRRHQFRWSIQGV-----RVTEPLNLPNPKRP--RRQWDGELYNGSFY 218
Db 124 NALISV-----SECDNKLKAFVNCVGLAGINDYEPFMPKLPKTYMSNGAIY 175
QY 219 F--AKRHLEMGYLOQGGKWHHTTCKELESHVDIIDWPIAEQ 259
Db 176 ILKIKEFLANNPSPQSKTKHFLMDE--SSSIDIDCLELKAQ 217

RESULT 8
US-09-328-352-6372
; Sequence 6372, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6372
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6372

Query Match 7.0%; Score 159; DB 4; Length 180;
Best Local Similarity 23.4%; Pred. No. 3.7e-08;
Matches 39; Conservative 48; Mismatches 70; Indels 10; Gaps 3;
QY 270 EKLKEIKLVNCDGCTNGHIYVSGDQKEIISYDVKDAIGISLKKSGIEVRLISERAC 329
Db 9 EQARHLQALVLDVGDGILSDGIVLINTGDEIKSFDROGLGMLKLAQQAQGMKVIITGRKS 68
QY 330 S--KQTLSSLLKLDCKMEVSVSKLAVVDWEKREMGKVCWEVAYLGNVEVSECKLKVGLS 387
Db 69 NIVEKRYSDLGVDLVPQ--GREDKGSALREACAQFNILPSDCLYMGDDWPDLSAFATAGMS 127
QY 388 GAPADACSAQKAQVYCKNGGGRGAIRFAEHIICL-----LMSEK 427

128 VTPNGHEVRRADLVTCQMGSGRGAVREVCNMLLIAGIYQELLEX 174

Best Local Similarity 22.4%; Pred No. 0.00028;
Matches 93; Conservative 59; Mismatches 154; Indels 110; Gaps 21;

US-08-867-611-28
Sequence 28, Application US/08867611
Patent No. 6,72189
GENERAL INFORMATION:
APPLICANT: DEVARE, SUSHIL G
APPLICANT: DESAI, SURESH M
APPLICANT: CASEY, JAMES M
APPLICANT: DAILEY, STEPHEN H
APPLICANT: DAWSON, GEORGE J
APPLICANT: GUTIERREZ, ROBIN A
APPLICANT: LESNIEWSKI, RICHARD R
APPLICANT: STEWART, JAMES L
APPLICANT: RUPPRECHT, KEVIN R
TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT ANTIGENS
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,611
FILING DATE: 02-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/646,757
FILING DATE:
APPLICATION NUMBER: US/08/179,896
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/572,822
FILING DATE: 24-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/614,069
FILING DATE: 07-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/748,561
FILING DATE: 21-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/748,565
FILING DATE: 21-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/748,566
FILING DATE: 21-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: FORENIESKI, PRISCILLA E
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834.US.P6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-867-611-28

Query Match 5.6%; Score 128; DB 3; Length 467;

QY 48 LILARGSGKIGPLKNIKHLGVLRAALDSGAFOSVWVSTHDHDIENVAKQFGAQ 107
DB 6 IIPARYASTRLPGKPLVDINGKXVIVHLERARESGA-ERIIIVATDHDVAAVEAAGGE 64
QY 108 Y-----HRRSEVSKDSSTSLDAIEELNYXNEXDIVCNICATSKLHPTDLQKVAEMI 161
DB 65 VCMTRADHQSGTE-----RLAEVVEKCAFSDDTTIV-VNQGDEPMIFATIIIRQVADNL 116
QY 162 REEGYDSXFSVVRHQFWSHIQKGVREVT-----PLNLNPAKRPRRCQWGDGELYENG-- 215
DB 117 AQRQVGMITLAVPIHNAEAFNPNAVKVLDAGGVALYFSRATIP-----WDRDRFAEGLE 172
QY 216 -----SFYFAKHLIEMGYLQ-----GKWHHTKCLEHSVDIDVDIADIAQORVLRGY 266
DB 173 TVGDNFL---RELGIYGYRAGFIRRYVMQPS--PLEH---IEM-----LEQLRLVWYG- 218
QY 267 FGKBEKLKEIKLLVCNIDGLTNGHIVSG-----DQKEIISYDVKDAIG-----IS 312
DB 219 ---EXI-----HVAHQEVPTGTVDPEDLDPSNTMGTGHPCTINTY 258
QY 313 LILK---KSGIEVRLISERACSKOTLSSILKLCNMEVSVSKLAVVDEWR-----KEMG 362
DB 259 LFKVEMVYGVGVEHRL--EVACNWTGERCDLDRDRSELSPDLLSTTQVLPSCFTTLP 316
QY 363 LCKKEVAYLGNVSDDECKRVGLSGAPADACSIAQK-----AVGVICKC 407
DB 317 ALTTGLIHLHQNIVDVQYLYGVGSS-----IVSWAKWEYVILLFILLADARICSC 367

RESULT 10
US-09-690-359-28
Sequence 28, Application US/09690359
Patent No. 6593083
GENERAL INFORMATION:
APPLICANT: DEVARE, SUSHIL G
APPLICANT: DESAI, SURESH M
APPLICANT: CASEY, JAMES M
APPLICANT: DAILEY, STEPHEN H
APPLICANT: DAWSON, GEORGE J
APPLICANT: GUTIERREZ, ROBIN A
APPLICANT: LESNIEWSKI, RICHARD R
APPLICANT: STEWART, JAMES L
APPLICANT: RUPPRECHT, KEVIN R
TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT ANTIGENS
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/690,359
FILING DATE: 17-Oct-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/867,611
FILING DATE: 02-JUN-1997
APPLICATION NUMBER: US/08/646,757
FILING DATE: <Unknown>
APPLICATION NUMBER: US/08/179,896
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/572,822

FILING DATE: 24-AUG-1990
 APPLICATION NUMBER: US 07/614,569
 FILING DATE: 07-NOV-1990
 APPLICATION NUMBER: US 07/748,561
 FILING DATE: 21-AUG-1991
 APPLICATION NUMBER: US 07/748,565
 FILING DATE: 21-AUG-1991
 APPLICATION NUMBER: US 07/748,566
 FILING DATE: 21-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: FOREMSKI, PRISCILLA E.
 REGISTRATION NUMBER: 33,207
 REFERENCE/DOCKET NUMBER: 4834.US.P6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 708-937-6365
 TELEFAX: 708-937-9556
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 467 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Query Match 5.6%; Score 128; DB 4; Length 467;
 Best Local Similarity 22.4%; Pred. No. 0.00028;
 Matches 93; Conservative 59; Mismatches 154; Indels 110; Gaps 21;

QY 48 LILARGSGKIPKNIKHLGAVPLIGWLRALDAGAFQSVWVSTHDEIENVAKQFGAQ 107
 DB 6 IIPARYASTPLPGKPLVDINGKPMIVHLERARESGA-ERIIATDHDVARAVEAAGGE 64

QY 108 V-----HRSSEVSKDSSLSDAIIIEFLNXXEXD:VGNIQATSCXCLHPTDLQKVAEMI 161
 DB 65 VCMTRADHQSGETE-----RLAEVVEKCAFSDDTV-V-VNQGDPEMPATIIRQVADNL 116

QY 162 REEGYDSXSFVRRHQFMRSEIQKGVREVT-----PLNLNPAKPRQDWDGELYENG-- 215
 DB 117 AQQVQKTTLAVPIHNAEAFNPAVKVLDAGYALYFSRATIP-----WDRDRFAEGLE 172

QY 216 ----SFYFAKRLHIEMGYLQ-----GKHTTKCELEHSDVIDDWPIAEQSVLYRGY 266
 DB 173 TVGDNPL----RHIGIVYGRAGFRRYVNWQPS--PLEH-----LEQRLVLYG- 218

QY 267 FGKELKEIKLLVCNIDGCLTNGHIYVSG-----DQKELISYDVKDAIG-----IS 312
 DB 219 ----EKI-----HVAQAQVPGTGVDTPELDLPSTNMTGHPCTINYT 258

QY 313 LLK-----KSGIEVRLISERACSKQTLLSKLCKMEVSVSDKLVAVDEWR-----KEMG 362
 DB 259 LFKVMYVGVGVEHRL--EVACNWTGRGCDLDRDRSELSPLLSTTQWVLPSCFTTLE 316

QY 363 LCKWEVAYLGNESVDECKRVGLSGAPADACSYAOK-----AVGYCKC 407
 DB 317 ALTTGLIHLHQNIVDVQYLYGVGSS-----IVSWAIKWEYVILLFLLLDARICSC 367

RESULT 11
 PCT-US92-06965A-33
 Sequence 33, Application PC/TUS9206965A
 GENERAL INFORMATION:
 APPLICANT: DEVARE, S.
 APPLICANT: DAILEY, S.
 TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
 NUMBER OF SEQUENCES: 35
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: ABBOTT LABORATORIES
 STREET: ONE ABBOTT PARK ROAD
 CITY: ABBOTT PARK
 STATE: ILLINOIS
 COUNTRY: U.S.

ZIP: 60065-3500
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/06965A
 FILING DATE: 19920821
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: FOREMSKI, PRISCILLA E.
 REGISTRATION NUMBER: 33,207
 REFERENCE/DOCKET NUMBER: 4834PC.02
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 708-937-6365
 TELEFAX: 708-937-9556
 INFORMATION FOR SEQ ID NO: 33:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 467 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US92-06965A-33

Query Match 5.6%; Score 128; DB 5; Length 467;
 Best Local Similarity 22.4%; Pred. No. 0.00028;
 Matches 93; Conservative 59; Mismatches 154; Indels 110; Gaps 21;

QY 48 LILARGSGKIPKNIKHLGAVPLIGWLRALDAGAFQSVWVSTHDEIENVAKQFGAQ 107
 DB 6 IIPARYASTPLPGKPLVDINGKPMIVHLERARESGA-ERIIATDHDVARAVEAAGGE 64

QY 108 V-----HRSSEVSKDSSLSDAIIIEFLNXXEXD:VGNIQATSCXCLHPTDLQKVAEMI 161
 DB 65 VCMTRADHQSGETE-----RLAEVVEKCAFSDDTVIV-VNQGDPEMPATIIRQVADNL 116

QY 162 REEGYDSXSFVRRHQFMRSEIQKGVREVT-----PLNLNPAKPRQDWDGELYENG-- 215
 DB 117 AQQVQKTTLAVPIHNAEAFNPAVKVLDAGYALYFSRATIP-----WDRDRFAEGLE 172

QY 216 ----SFYFAKRLHIEMGYLQ-----GKHTTKCELEHSDVIDDWPIAEQSVLYRGY 266
 DB 173 TVGDNPL----RHIGIVYGRAGFRRYVNWQPS--PLEH-----LEQRLVLYG- 218

QY 267 FGKELKEIKLLVCNIDGCLTNGHIYVSG-----DQKELISYDVKDAIG-----IS 312
 DB 219 ----EKI-----HVAQAQVPGTGVDTPELDLPSTNMTGHPCTINYT 258

QY 313 LLK-----KSGIEVRLISERACSKQTLLSKLCKMEVSVSDKLVAVDEWR-----KEMG 362
 DB 259 LFKVMYVGVGVEHRL--EVACNWTGRGCDLDRDRSELSPLLSTTQWVLPSCFTTLE 316

QY 363 LCKWEVAYLGNESVDECKRVGLSGAPADACSYAOK-----AVGYCKC 407
 DB 317 ALTTGLIHLHQNIVDVQYLYGVGSS-----IVSWAIKWEYVILLFLLLDARICSC 367

RESULT 12
 US-08-867-611-8
 Sequence 8, Application US/98867511
 Patent No. 6172189
 GENERAL INFORMATION:
 APPLICANT: DEVARE, SUSHIL G
 APPLICANT: DESAI, SURESH M
 APPLICANT: CASSEY, JAMES M
 APPLICANT: DAILEY, STEPHEN H
 APPLICANT: DAWSON, GEORGE J
 APPLICANT: GUTIERREZ, ROBIN A
 APPLICANT: LESNIEWSKI, RICHARD R
 APPLICANT: STEWART, JAMES L
 APPLICANT: RUPPRECHT, KEVIN R
 TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT

US-09-690-359-8

Query Match 5.6%; Score 128; DB 4; Length 541;
Best Local Similarity 22.7%; Pred. No. 0.00035;
Matches 78; Conservative 56; Mismatches 131; Indels 78; Gaps 17;
QY 48 LILARGSGKIPKNIKHLGVLPLIGWLRALDSGAFQSVWVSTHDEIENVAQFGAQ 107
DB 6 IIPARYASTRLPGKPLVDINGKPMIVHLEARESGA-ERIVATDHEDVARAVEAAGE 64
QY 108 V-----HRSSEVSKDSSLSDAIIEFLNYXNEXDIVGNIOATSCXCLHPTDLQKVAEMI 161
DB 65 VCMTRADHQSGTE-----RLAEVVEKCAFSDDTIVV-NVQGDPEMPATIIQVADNL 116
QY 162 REEYDSXFSVRRHQRWSEIQKGVREVT-----PLNLNPAKPRRQDWDGELYENG-- 215
DB 117 AQRQVGMATLAVPHHNAEEAFNPNVAVLDAGYALYFSRATIP-----WDRORFAEGLE 172
QY 216 ----SYFAKRLHLEMGYLOG-----GKWHITKCELEHSVDIDVDIPIAEOQVLRVGY 266
DB 173 TVGDNFL---RHIGIYGRAGFIRRYVNWQPS---PLEH---IEM-----LEQLRVLMYG- 218
QY 267 EGKELKEIKLVNIDGCLTNGHYVSGQKEIISYDVK-----DAIGISLLKXSGIEV 321
DB 219 ---EKI-----HVAVA---QEVPGTGVDTPELDLPSTNSMRRLARGSP 255
QY 322 RLISERACSKQTLSSKLCKMEVSVSDKLAVYDE--WRXEMG 362
DB 256 PSVASSASQLSAPSLKATCTANHSDPAELIEANLLWRQEMG 298

RESULT 14

PCT-US92-06965A-13
Sequence 13, Application PC/TUS9206965A
GENERAL INFORMATION:
APPLICANT: DEVARE, S.
APPLICANT: DESAI, S.
APPLICANT: DAILEY, S.
TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: ILLINOIS
COUNTRY: U.S.
ZIP: 60065-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06965A
FILING DATE: 19920821
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834PC.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 541 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-06965A-13

Query Match 5.6%; Score 128; DB 5; Length 541;
Best Local Similarity 22.7%; Pred. No. 0.00035;

Matches 78; Conservative 56; Mismatches 131; Indels 78; Gaps 17;

QY 48 LILARGSGKIPKNIKHLGVLPLIGWLRALDSGAFQSVWVSTHDEIENVAQFGAQ 107
DB 6 IIPARYASTRLPGKPLVDINGKPMIVHLEARESGA-ERIVATDHEDVARAVEAAGE 64
QY 108 V-----HRSSEVSKDSSLSDAIIEFLNYXNEXDIVGNIOATSCXCLHPTDLQKVAEMI 161
DB 65 VCMTRADHQSGTE-----RLAEVVEKCAFSDDTIVV-NVQGDPEMPATIIQVADNL 116
QY 162 REEYDSXFSVRRHQRWSEIQKGVREVT-----PLNLNPAKPRRQDWDGELYENG-- 215
DB 117 AQRQVGMATLAVPHHNAEEAFNPNVAVLDAGYALYFSRATIP-----WDRORFAEGLE 172
QY 216 ----SYFAKRLHLEMGYLOG-----GKWHITKCELEHSVDIDVDIPIAEOQVLRVGY 266
DB 173 TVGDNFL---RHIGIYGRAGFIRRYVNWQPS---PLEH---IEM-----LEQLRVLMYG- 218
QY 267 EGKELKEIKLVNIDGCLTNGHYVSGQKEIISYDVK-----DAIGISLLKXSGIEV 321
DB 219 ---EKI-----HVAVA---QEVPGTGVDTPELDLPSTNSMRRLARGSP 255
QY 322 RLISERACSKQTLSSKLCKMEVSVSDKLAVYDE--WRXEMG 362
DB 256 PSVASSASQLSAPSLKATCTANHSDPAELIEANLLWRQEMG 298

RESULT 15
US-08-867-611-48
Sequence 48, Application US/08867611
Patent No. 6172189
GENERAL INFORMATION:
APPLICANT: DEVARE, SUSHIL G
APPLICANT: DESAI, SURESH M
APPLICANT: CASEY, JAMES M
APPLICANT: DAILEY, STEPHEN H
APPLICANT: DAWSON, GEORGE J
APPLICANT: GUTIERREZ, ROBIN A
APPLICANT: LESNIEWSKI, RICHARD R
APPLICANT: STEWART, JAMES L
APPLICANT: RUPPRECHT, KEVIN R
TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,611
FILING DATE: 02-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/646,757
FILING DATE:
APPLICATION NUMBER: US/08/179,896
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/572,822
FILING DATE: 24-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/614,069
FILING DATE: 07-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,561

```

: FILING DATE: 21-AUG-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/748,565
: FILING DATE: 21-AUG-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/748,566
: FILING DATE: 21-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: PCREMBSKI, PRISCILLA E
: REGISTRATION NUMBER: 33,207
: REFERENCE/DOCKET NUMBER: 4834.US.P6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 708-937-6365
: TELEFAX: 708-937-9556
: INFORMATION FOR SEQ ID NO: 48:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 594 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Peptide
: US-08-867-611-48

Query Match: 5.6%; Score 128; DB 3; Length 594;
Best Local Similarity 22.2%; Pred. No. 0.0041;
Matches 99; Conservative 61; Mismatches 141; Indels 144; Gaps 23;

QY 48 LILARGSKGKPLKNIKHLGVLGVLGALDGAQSQSVVSTDDHDEIENVAQKQFGAQ 107
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6 IIPARYASTRLPGKPLVDINGKPMIVHVLIERASGA-ERIIVATDHDVAVAEAAAGE 64
QY 108 V-----HRRSEVSKDSLSDAIIEFLNYXNEXDIVGNIQATSKLHPTDLQKVAEMI 161
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
65 VCNTPADHQSGTE-----RLAEVVEKAFSDDTIVV-NVQDEPMIPATIIHQVADKL 116
QY 162 REEGYDSKFSVVRHQFRWSEIQKQVRYTE-----PLNLNPAKRPRRQDWDGELYENG-- 215
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
117 AQROVGMATLAVPIHNABEAFNPNAVKVVDAGVVALYFSRATIP---WEDRFAEGLE 172
QY 216 ----SFYAKRHLIEMGYLOG-----GKWHITKCELEHSVDIDVDWPIAEQRLVRYGY 266
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
173 TWGDNFL---RHLGIYGVRAGFIRYVNWQPS--PLEH-----LEQLRVILWYG- 218
QY 267 PKEKELKEIKLIVCNIDGCLINGHIYVSG-----DQKEII--SYLVKDAIGISLLKKS 317
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
219 ---EKI-----HVAAQEVPGTGVDPEDLDPTNSMDAHFLSQTKGS 258
QY 318 SIEV-RLISERA--CSKQ-----TLSSKLDCQKQEVSVSDKLVVDEN 357
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
259 SENLYLVAYQATVCARAQAAPTLLYRLGAVQNEITLTHPTVKYIMTCMSADLEVVTSTW 318
QY 358 RKEMGLCWKEVAYLCNEVSDSECLK-----RVCLSGAPA-----DAC 394
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
319 VLVGGVLAALANY-----CLSTGCVVIYGRVLSGKPALIIPDREVLYREFDEMEEC 369
QY 395 S-----YAKQAVGYI 404
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
370 SOHLPIYEQXVMLEAFQFKQKALGLL 394

```

Search completed: May 6, 2004, 09:08:36
Job time : 19.0701 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:07:45 ; Search time 46.9189 Seconds
(without alignments)
2567.492 Million cell updates/sec

Title: US-09-930-440B-4

Perfect score: 2266

Sequence: 1 MDSVERGATEVSNPRGRPS.....REFAEHICLLMEXWNNSCQK 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2258	99.6	434	9	US-09-984-205-2
2	2258	99.6	434	9	US-09-930-440B-4
3	2188	95.6	448	9	US-09-925-301-1376
4	1173	51.8	245	15	US-10-012-697-1535
5	945	41.7	202	9	US-09-867-550-1194
6	376	16.6	432	14	US-10-156-761-10911
7	339	15.0	74	11	US-09-864-408A-3298
8	241	10.6	44	9	US-09-864-761-35760
9	225.5	10.0	169	9	US-09-870-162A-7
10	199.5	9.2	410	9	US-09-767-041-28
11	139.5	8.8	164	12	US-10-335-977-7689
12	177.5	7.8	209	12	US-10-282-122A-61379
13	169.5	7.5	221	9	US-09-816-028A-37
14	169.5	7.5	221	14	US-10-303-161-37
15	169.5	7.5	221	14	US-10-303-118-37

16	169.5	7.5	221	14	US-10-303-128-37	Sequence 37, Appl
17	169.5	7.5	221	14	US-10-303-134-37	Sequence 37, Appl
18	169.5	7.5	221	14	US-10-303-162-37	Sequence 37, Appl
19	145	6.4	536	9	US-09-816-028A-19	Sequence 19, Appl
20	145	6.4	536	9	US-09-816-028A-25	Sequence 25, Appl
21	145	6.4	536	14	US-10-303-161-19	Sequence 19, Appl
22	145	6.4	536	14	US-10-303-161-25	Sequence 25, Appl
23	145	6.4	536	14	US-10-303-118-19	Sequence 19, Appl
24	145	6.4	536	14	US-10-303-118-25	Sequence 25, Appl
25	145	6.4	536	14	US-10-303-128-19	Sequence 19, Appl
26	145	6.4	536	14	US-10-303-128-25	Sequence 25, Appl
27	145	6.4	536	14	US-10-303-134-19	Sequence 19, Appl
28	145	6.4	536	14	US-10-303-134-25	Sequence 25, Appl
29	145	6.4	536	14	US-10-303-162-19	Sequence 19, Appl
30	145	6.4	536	14	US-10-303-162-25	Sequence 25, Appl
31	142.5	6.3	229	12	US-10-335-977-6838	Sequence 60, Appl
32	120	5.3	460	8	US-08-911-824-60	Sequence 55, Appl
33	120	5.3	466	8	US-08-911-824-55	Sequence 50, Appl
34	120	5.3	490	8	US-08-911-824-50	Sequence 97, Appl
35	120	5.3	526	8	US-08-911-824-97	Sequence 54, Appl
36	120	5.3	618	8	US-08-911-824-54	Sequence 56, Appl
37	119.5	5.3	491	8	US-08-911-824-56	Sequence 108, Appl
38	119.5	5.3	599	8	US-08-911-824-108	Sequence 93, Appl
39	119.5	5.3	706	8	US-08-911-824-93	Sequence 47912, A
40	119.5	5.3	736	8	US-08-911-824-91	Sequence 6837, A
41	115	5.1	31	9	US-09-864-761-47912	Sequence 7688, A
42	110	4.9	99	12	US-10-335-977-6837	Sequence 54, Appl
43	109.5	4.8	90	12	US-10-335-977-7688	Sequence 54, Appl
44	109.5	4.8	398	9	US-09-896-852-54	Sequence 54, Appl
45	109.5	4.8	398	10	US-09-728-644-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1
US-09-984-205-2
; Sequence 2, Application US/09984205
; Patent No. US20020137175A1
; GENERAL INFORMATION:
; APPLICANT: Coleman, Timothy A. et al.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: EP0505D1
; CURRENT APPLICATION NUMBER: US/09/984,205
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: PCT/US00/05325
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/516,143
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 133
; OTHER INFORMATION: Xaa equals Tyr or His
; NAME/KEY: SITE
; LOCATION: 136
; OTHER INFORMATION: Xaa equals Gly or Val
; NAME/KEY: SITE
; LOCATION: 147
; OTHER INFORMATION: Xaa equals Ser or Pro
; NAME/KEY: SITE
; LOCATION: 169
; OTHER INFORMATION: Xaa equals Gly or Val
US-09-984-205-2

Query Match

99.6%; Score 2258; DB 9; Length 434;


```
Best Local Similarity 100.0%; Pred. No. 5.5e-220;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVSNPRGPRSPKLRNRRGGQGRGVEKPPHLLAALILARGSGKGIPL 60
DB 1 MDSVEKGAATSVSNPRGPRSPKLRNRRGGQGRGVEKPPHLLAALILARGSGKGIPL 60
QY 61 KNIKHLAGVPLIGWVLRRAALDSGAFQSVVWSTDHDEIENVAQFGAQVHRRSRSEVSKDSS 120
DB 61 KNIKHLAGVPLIGWVLRRAALDSGAFQSVVWSTDHDEIENVAQFGAQVHRRSRSEVSKDSS 120
QY 121 TSLDAIIEFLNXXEXDIVGNQIATSCXCLHPTDLQKVAEMIREEGYDSFVSFVRRHQFHW 180
DB 121 TSLDAIIEFLNXXEXDIVGNQIATSCXCLHPTDLQKVAEMIREEGYDSFVSFVRRHQFHW 180
QY 181 SEIQGVRETEPLNPAKPRRQDWDGELYENGSGFYPAKRLHTEMGLQGGKWHHTKC 240
DB 181 SEIQGVRETEPLNPAKPRRQDWDGELYENGSGFYPAKRLHTEMGLQGGKWHHTKC 240
QY 241 ELEHSDVDIDWDPIAEQVRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGQKEI 300
DB 241 ELEHSDVDIDWDPIAEQVRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGQKEI 300
QY 301 ISYDVKDAIGISLLKXKSGIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDWEWKE 360
DB 301 ISYDVKDAIGISLLKXKSGIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDWEWKE 360
QY 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQKAVGYICKNGRGRAIREFAEH 420
DB 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQKAVGYICKNGRGRAIREFAEH 420
QY 421 ICLLMEKVNNSCQK 434
DB 421 ICLLMEKVNNSCQK 434

RESULT 2
US-09-930-440b-4
; Sequence 4, Application US/09930440B
; Patent No. US20020142386A1
; GENERAL INFORMATION:
; APPLICANT: Betenbaugh et al.
; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
; FILE REFERENCE: PF509P2
; CURRENT APPLICATION NUMBER: US/09/930,440B
; CURRENT FILING DATE: 2001-08-16
; PRIOR FILING DATE: 60/227,579
; PRIOR FILING DATE: 2000-08-25
; PRIOR FILING DATE: 09/516,793
; PRIOR FILING DATE: 2000-03-01
; PRIOR FILING DATE: 60/169,624
; PRIOR FILING DATE: 1999-12-08
; PRIOR FILING DATE: 60/122,582
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (133)..(133)
; OTHER INFORMATION: The 'Xaa' at location 133 stands for His, or Tyr.
; NAME/KEY: misc feature
; LOCATION: (136)..(136)
; OTHER INFORMATION: The 'Xaa' at location 136 stands for Gly, or Val.
; NAME/KEY: misc feature
; LOCATION: (147)..(147)
; OTHER INFORMATION: The 'Xaa' at location 147 stands for Pro, or Ser.
; NAME/KEY: misc feature
; LOCATION: (169)..(169)
; OTHER INFORMATION: The 'Xaa' at location 169 stands for Gly, or Val.

US-09-930-440b-4
Query Match 99.6%; Score 2188; DB 9; Length 434;
Best Local Similarity 100.0%; Pred. No. 5.5e-220;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVSNPRGPRSPKLRNRRGGQGRGVEKPPHLLAALILARGSGKGIPL 60
DB 1 MDSVEKGAATSVSNPRGPRSPKLRNRRGGQGRGVEKPPHLLAALILARGSGKGIPL 60
QY 61 KNIKHLAGVPLIGWVLRRAALDSGAFQSVVWSTDHDEIENVAQFGAQVHRRSRSEVSKDSS 120
DB 61 KNIKHLAGVPLIGWVLRRAALDSGAFQSVVWSTDHDEIENVAQFGAQVHRRSRSEVSKDSS 120
QY 121 TSLDAIIEFLNXXEXDIVGNQIATSCXCLHPTDLQKVAEMIREEGYDSFVSFVRRHQFHW 180
DB 121 TSLDAIIEFLNXXEXDIVGNQIATSCXCLHPTDLQKVAEMIREEGYDSFVSFVRRHQFHW 180
QY 181 SEIQGVRETEPLNPAKPRRQDWDGELYENGSGFYPAKRLHTEMGLQGGKWHHTKC 240
DB 181 SEIQGVRETEPLNPAKPRRQDWDGELYENGSGFYPAKRLHTEMGLQGGKWHHTKC 240
QY 241 ELEHSDVDIDWDPIAEQVRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGQKEI 300
DB 241 ELEHSDVDIDWDPIAEQVRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGQKEI 300
QY 301 ISYDVKDAIGISLLKXKSGIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDWEWKE 360
DB 301 ISYDVKDAIGISLLKXKSGIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDWEWKE 360
QY 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQKAVGYICKNGRGRAIREFAEH 420
DB 361 MGLCWKEVAYLGNESVDEECLKRVGLSGAPADACSYAQKAVGYICKNGRGRAIREFAEH 420
QY 421 ICLLMEKVNNSCQK 434
DB 421 ICLLMEKVNNSCQK 434

RESULT 3
US-09-925-301-1376
; Sequence 1376, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR FILING DATE: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1376
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (137)..(137)
; OTHER INFORMATION: The 'Xaa' at location 137 stands for His, or Tyr.
; NAME/KEY: misc feature
; LOCATION: (147)..(147)
; OTHER INFORMATION: The 'Xaa' at location 147 stands for Pro, or Ser.
; NAME/KEY: misc feature
; LOCATION: (169)..(169)
; OTHER INFORMATION: The 'Xaa' at location 169 stands for Gly, or Val.
```

QY 121 TSLDAIEFLNLYNEXDIVGNIOATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180
DB 135 TSLDAIEFLNLYNEXDIVGNIOATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 194
QY 181 SEIQKGVRETEPLNPAKRPRQDMDGSLYENGSGFYFAKRLHLEMGYLOGGKWHHTKC 240
DB 195 SEIQKGVRETEPLNPAKRPRQDMDGSLYENGSGFYFAKRLHLEMGYLOGGKWHHTKC 254
QY 241 ELESHVDIDVDIPIAEQRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
DB 255 RAHSHVDIDVDIPIAEQRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGDQKEI 314
QY 301 ISYDVDAAGISLLKSGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLAUVDEWRKE 360
DB 315 ISYDVDAAGISLLKSGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLAUVDEWRKE 374
QY 361 MGLCWEVAYLNGEVSDDEELCKRVGLSGAPADACSVAKVAGVYCKNGGGRGAIRFAEH 420
DB 375 MGLCWEVAYLNGEVSDDEELCKRVGLSGAPADACSVAKVAGVYCKNGGGRGAIRFAEH 434
QY 421 ICLLMKXVNSCQK 434
DB 435 ICLLMKXVNSCQK 448

RESULT 4
US-10-012-697-1535
; Sequence 1535, Application US/10012697
; Publication No. US20030215803A1
; GENERAL INFORMATION:
; APPLICANT: Escobedo, Jaime
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Scott, Beth
; APPLICANT: Drmanac, Radoje
; APPLICANT: Cirkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS
; FILE REFERENCE: 2300-16252
; CURRENT APPLICATION NUMBER: US/10/012,697
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/254,648
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 60/275,668
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 1568
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1535
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-697-1535

Query Match 51.8%; Score 1173; DB 15; Length 245;
Best Local Similarity 95.7%; Pred. No. 3.1e-110;
Matches 225; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 137 DIVGNIOATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRWSIOLKGVRETEPLN 196
DB 3 DIVGNIOATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRWSIOLKGVRETEPLN 62
QY 197 NPAKRPRQDMDGSLYENGSGFYFAKRLHLEMGYLOGGKWHHTKCLEHSHVDIDVDIPI 256
DB 63 NPAKRPRQDMDGSLYENGSGFYFAKRLHLEMGYLOGGKWHHTKCLEHSHVDIDVDIPI 122

QY 257 AEQRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGDQKEIISYDVDAAGISLLK 316
DB 123 AEQRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGDQKEIISYDVDAAGISLLK 182
QY 317 SGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLAUVDEWRKEMGLCWEVAYL 371
DB 183 SGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLAUVDEWRKEMGLCWEVAYL 237

RESULT 5
US-09-867-550-1194
; Sequence 1194, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1194
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (7)
; OTHER INFORMATION: Wherein Xaa may be any one of Arg or Gly or Trp
; NAME/KEY: VARIANT
; LOCATION: (12)
; OTHER INFORMATION: Wherein Xaa may be any one of Ala or Arg or Gln or Glu or Gly or
; OTHER INFORMATION: Lys or Met or Pro or Ser or Thr or Trp or Val
; NAME/KEY: VARIANT
; LOCATION: (41)
; OTHER INFORMATION: Wherein Xaa may be any one of Ala or Pro or Ser or Thr
US-09-867-550-1194

Query Match 41.7%; Score 945; DB 9; Length 202;
Best Local Similarity 92.2%; Pred. No. 3.2e-87;
Matches 188; Conservative 1; Mismatches 13; Indels 2; Gaps 1;
QY 1 MDSVEKGAATSVSNRGRPSRGRPPKLRNSRGGQGRGVEKPPHLLAALLIARGSGKIPL 60
DB 1 MDSVEKGAATSVSNRGRPSRGRPPKLRNSRGGQGRGVEKPPHLLAALLIARGSGKIPL 58
QY 61 KNKHLAGVPLIGWVLRALDSCAFOSVWVSTHDEIENVAKQFGAQVHRSSEVSKDS 120
DB 59 KNKHLAGVPLIGWVLRALDSCAFOSVWVSTHDEIENVAKQFGAQVHRSSEVSKDS 118
QY 121 TSLDAIEFLNLYNEXDIVGNIOATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180
DB 119 TSLDAIEFLNLYNEXDIVGNIOATSKCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 178
QY 181 SEIQKGVRETEPLNPAKRPRR 204
DB 179 SEIQKGVRETEPLNPAKRPRR 202

RESULT 6
US-10-156-761-10911
; Sequence 10911, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI

```
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10911
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10911

Query Match      16.6%; Score 376; DB 14; Length 432;
Best Local Similarity 27.4%; Pred. No. 6.1e-29;
Matches 118; Conservative 75; Mismatches 188; Indels 50; Gaps 9;

Qy      30 NSRGQGRGVEKPHLAALILARGSGKGIPLKNIKHLAGVPLIGWLVRAALDSGAFQSVW 89
Db      3 HSEAGQAASVTR--VLAIVFARGSGKVPKXNLLFVGVPVVARARECATRLVTDVV 59

Qy      90 VSTDHDEIENAVAKQFGAQVHRSEVSKDSSTSLDAIEFILN-----YXNEXDIIVGNIQA 144
Db      60 VSTDHDAIAAAAREAGAEVVLRLPAALAGDTATSEAAVLHAMDTEALHGAADVLLVQC 119

Qy      145 TSXCLHPTDLQVAMIEEESYDSVSVVRHQRWSEI-----QKGVREVIETPL 194
Db      120 TSPFIVREDVDGVVHAIIVGKADTALTVAPEHGFVWRDADAPGALGAERAAVEGGDTL 179

Qy      195 -----NLNPAKPRQGWGELYENGSPYFAKRLHBMGYLQG-----GKWH 237
Db      180 VTGATSGGYGNHDKSFRPRQRPQDLLETGAVY-----GKATGTFREAKHRFGRETEL 235

Qy      238 TKCELEHSVDIDVDIDMPIAEQRLV-----RYGVFGKEKLEKLEKLVNIDGCLNGH 290
Db      236 VRTDPAKVLIEDPHD--LARAALAPLFDANRPG--ALPTAESIDAVVLDGDTQDTR 291

Qy      291 IYVSGDQKEI--SYDVKDAIGISLLKSGIEVRLISERACSKQTLSLKLCKMEVSUSDK 350
Db      292 VLIDSDDGREFSVHRGDDGLGAALRKSGLTMLILSTEQNPVVAARAKLKIPVLHGIDRK 351

Qy      351 LAVYDEVRKEMGLCWKEVAYLGNESDEECLKRVGLSGAPADACSYAQKAVGYICKNGG 410
Db      352 DLALKQWCEOGIAPERVLYVGNVDNDLPCFALYGVKPVAVASAEVDVVGARAVTTVPGG 411

Qy      411 RGATREFAEHI 421
Db      412 DGAIRETASWI 422

RESULT 7
US-09-864-408A-3298
; Sequence 3298, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Shimkars, Richard A.
; TITLE OF INVENTION: NO. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35760
; LENGTH: 44

; SEQ ID NO 3298
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-3298

Query Match      15.0%; Score 339; DB 11; Length 74;
Best Local Similarity 86.3%; Pred. No. 2.2e-26;
Matches 63; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy      204 RODWDGELYENGSPYFAKRLHBMGYLQGGKWHHTTKCELEHSVDIDVDIDMPIAEQRLV 263
Db      1 RODWDGELYENGSPYFAKRLHBMGYLQGGKWHHTTKCELEHSVDIDVDIDMPIAEQRLV 60

Qy      264 YGYFGKEKLEK 276
Db      61 FGYFGKEKLEK 73

RESULT 8
US-09-864-761-35760
; Sequence 35760, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35760
; LENGTH: 44
```

; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC007671.5
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.99
 ; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.1
 ; OTHER INFORMATION: SWISSPROT HIT: P38144, EVALUOR 6.13e+00
 ; OTHER INFORMATION: EST_HUMAN HIT: BE729942.1, EVALUOR 2.00e-20
 ; US-09-864-761-35760

Query Match 10.6%; Score 241; DB 9; Length 44;
 Best Local Similarity 100.0%; Pred. No. 8.7e-17;
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 REVTEPLNLPAPRRQDMDGELYENGSEFYFAKRLHLEMGYLQ 231
 Db 1 REVTEPLNLPAPRRQDMDGELYENGSEFYFAKRLHLEMGYLQ 44

RESULT 9
 US-09-870-162A-7
 ; Sequence 7, Application US/09870162A
 ; Patent No. US20020042118A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Breinig, Sabine
 ; APPLICANT: Fuchs, Georg
 ; TITLE OF INVENTION: Phenol Induced Proteins of Thauera aromatica
 ; FILE REFERENCE: BC1006 US DIV
 ; CURRENT APPLICATION NUMBER: US/39/870,162A
 ; CURRENT FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: 09/516914
 ; PRIOR FILING DATE: 2000-03-01
 ; NUMBER OF SEQ ID NOS: 44
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 7
 ; LENGTH: 169
 ; TYPE: PRT
 ; ORGANISM: Thauera aromatica
 ; US-09-870-162A-7

Query Match 10.0%; Score 225.5; DB 9; Length 169;
 Best Local Similarity 31.2%; Pred. No. 2.7e-14;
 Matches 48; Conservative 39; Mismatches 64; Indels 3; Gaps 2;

QY 270 EKUKELKLVNIDGCTNGHIYVSGQKEIISVDVKAIGISLLKSGIEVRLISER-- 327
 Db 2 EQAKNIKLVLDVGVWTDGRIVNDEGIESRNFIDKGMGVIVLQCGVEVAITSKS 61
 QY 328 ACSQTLSSLLKLDCKMEVSVSDKLAVDENEKMGKLCWEKAVYLGNEVSDEECLKRYGLS 387
 Db 62 GAVHRAEELKIK-RFHEGIKKKTPEYQMLDEWNISDAEVCYVGBDLVLSMKRVGLA 120

RESULT 10
 US-09-767-041-28
 ; Sequence 28, Application US/09767041
 ; Patent No. US20020055168A
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Hilda
 ; TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
 ; FILE REFERENCE: 2183-4726
 ; CURRENT APPLICATION NUMBER: US/09/767,041
 ; CURRENT FILING DATE: 2001-01-22

; PRIOR APPLICATION NUMBER: PCT/NL99/0046C
 ; PRIOR FILING DATE: 1999-07-19
 ; PRIOR APPLICATION NUMBER: EP98202465.5
 ; PRIOR FILING DATE: 1998-07-22
 ; PRIOR APPLICATION NUMBER: EP98202467.1
 ; PRIOR FILING DATE: 1998-07-22
 ; NUMBER OF SEQ ID NOS: 53
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 28
 ; LENGTH: 410
 ; TYPE: PRT
 ; ORGANISM: Streptococcus suis
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: CPS2T
 ; US-09-767-041-28

Query Match 9.2%; Score 209.5; DB 9; Length 410;
 Best Local Similarity 23.5%; Pred. No. 4.5e-12;
 Matches 93; Conservative 71; Mismatches 149; Indels 83; Gaps 14;

QY 48 LILARGSGKGPLKNKHLKLAGVPLIGWVIRALDLSGAF--OSVWVSTHDELENVAKQFG 105
 Db 6 LIPARSGSKGLPNKNMLFLDGVPMIFHTIRAAIBSGCFKENIYVSTDSEVYKEICETTG 65
 QY 106 AOVHRSSEVSKDSTSLDAIEPLNYNEXDIVGNIQATSKLHPTLQKVAEMIREEG 165
 Db 66 VQVLMRPADLADFTTSPQLMEHFLQDPSDOOVFLLQVTSPLRSGKHVKZEMELYGKGQ 125
 QY 166 YDSXFSVVRHQFRWSHIQKGVRETEPLNLPAPRRP-----RRQDWDGELYENGSEF 217
 Db 126 AD-----HVSFTKVDKS-PTLFSTLDENGFAKOLAGLGSVRRQDEKLTLYPENGAI 176
 QY 218 YFAKHLIEMGYLOGSKWHHTKCEL-----BHSVDIDVDIDWPIAEQVRLRVGY----- 266
 Db 177 YISSK---QAYLADKTYFSEKTAAYVMTKEDSIDVDHFDFTGVIIGRIY-FDYQRREQQ 231
 QY 267 ---FGKEKLKEI-----KLVVCNIDG-----CLTNGHIYVSGD 296
 Db 232 NKPFYKRELKRLCQRVHDSLVIGDSRLALLLDGFDNISIGGTMSTASLENGQGLFATP 291
 QY 297 -QKEIISYDVVDKDAICISLLKSGIEVRLISERACSK-----QTLSSLLKLDCKMEV 345
 Db 292 IKKVLISLVGNDLITDVEHMIEDTIRQLMESLVSKAQEVVTTIAYTLFSDSVSNBEV 351
 QY 346 SVSKLAVVDEWRKEMGLCWKEVAYLGNEVSDEECL 381
 Db 352 QLND---VIVQASSELGIGVIDL---NEVVEKEAM 380

RESULT 11
 US-10-335-977-7689
 ; Sequence 7689, Application US/10335977
 ; Publication No. US20040052799A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGLAS SMITH et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; RELATING TO HELICOBACTER PYLORI FOR
 ; DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 10031
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 28 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-ROM ISO9660
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: Windows NT 4.0
 ; SOFTWARE: UNIX
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 7689:
SEQUENCE CHARACTERISTICS:
LENGTH: 164 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
OR-GINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...164
SEQUENCE DESCRIPTION: SEQ ID NO: 7689:
US-10-335-977-7689

Query Match 8.8%; Score 199.5; DB 12; Length 164;
Best Local Similarity 30.3%; Pred. No. 1.1e-11;
Matches 47; Conservative 43; Mismatches 50; Indels 15; Gaps 2;
QY 275 IKLVNIDGCTNGHIVYSGQKEIIISYDVKDAIGISILKKSGIEVRLISERACKQTL 334
DB 2 IKLLLDVDTLTDGSLYFDEFNHFIKAFNVKGLGMLTQWKLKKIATIGR----- 54
QY 335 SSKLDCRME-----VVSDDKLVVDENKEMGLCKWKEVAYLGNVSDDECKRVGL 386
DB 55 TSIMVKRMESLGQVFVGVGVKNVAVIERLKKDLQLSACEIACVGDYNDLGMFKACTL 114
QY 387 SGAPADACSYAQKAVGYICKNGGRCAGTREFAEHI 421
DB 115 SFAPDAPHLKSKAYKVLQNSGGKGVAREADYL 149

RESULT 12
US-10-282-122A-61379
Sequence 61379, Application US/10282122A
Publication No. US20040329129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Chlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELTRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 61379
LENGTH: 209
TYPE: PRT
ORGANISM: Legionella pneumophila
US-10-282-122A-61379
Query Match 7.8%; Score 177.5; DB 12; Length 209;
Best Local Similarity 29.8%; Pred. No. 2.8e-09;
Matches 61; Conservative 31; Mismatches 68; Indels 45; Gaps 6;
QY 45 LAALILARGSGKIPKNIKHLGVPLIGWLRALDGAFOVWVSTDHDEIENVAKOF 104
DB 3 ILAVIPARAGSKKLPCKNTRLLAGKELTAHTIVAAQLQSCCEIVVSTDKGIADVAVY 52
QY 105 GAQV-HRRSSEVKSQSTSLDAIEFLNYXNEXDI-----VGNQATSKLHPTDLQKVAE 159
DB 63 GASVPLWRSEDLATDTSVHTVIDLLEFKFOQMDVFDLSVLLQPTSPRKPETISHAVE 122
QY 160 M'REEG-----YDSFESVVRQFQWSEIQKGVRETEPLINLPAKE 201
DB 123 IHKVTGKSVVSPISLKPSCRSIDSQNLVKPLFQDLEY-----CNENPIYX 173
QY 202 PRQWDGELY-----ENGsfy 218
DB 174 -----LNGSIVATAKQIENKsfy 193

RESULT 13
US-09-816-028A-37
Sequence 37, Application US/09816028A
Patent No. US20020042369A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarchuk, Warren W.
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/09/816,028A
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 37
LENGTH: 221
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: CMP-sialic acid synthetase from C. jejuni OH4384
OTHER INFORMATION: (ORF 10a of lipooligosaccharide (LOS) biosynthesis
OTHER INFORMATION: locus)
US-09-816-028A-37
Query Match 7.5%; Score 169.5; DB 9; Length 221;
Best Local Similarity 28.3%; Pred. No. 2e-08;

	Matches	63;	Conservative	35;	Mismatches	136;	Indels	19;	Gaps	5;
Qy	47	ALILARGGSGIPLKNIKHLAGVPLICWLEAALDSGAFOSVWYSTDHDIEIRYAKOFGA	106							
Dd	4	AIIPARGGSGIKNNKLVLLNNKPLIYTTIKAAALNTKISKVVVSSDSDEIINYAKSONV	63							
Qy	107	QVHRRSSEVSKDSTSLDAIEFLINYXNEXDIVGNIQATSKXCLHPTDLQKVAEMIREEGY	165							
Dd	64	DILKRPISLAQDNTSTCKVLLHALFKYQYEDVVFLQTPSLPTNIIHDEAFNLYKNSNA	123							
Qy	167	DSXFSVVRHQFRWSEIQKGV-----REVTEPLNLNPAKRP--RRQWDGELYENGFSFY	218							
Dd	124	NALISV-----SECDNKLKAFVCEVGYDLAGINCDEYFEMPRQKLPTIYMSNGAIY	175							
Qy	219	F--AKRHLEFVGYLQGGKWHFTKCELSHSDVIDVDIDWPIAEQ	259							
Dd	176	ILKTKETFNPPSFLOSKTKFELMDE-SSSLIDICLEDLKKAEQ	217							

RESULT 14

```

RECORD 14
US-10-303-161-37
; Sequence 37, Application US/10303161
; Publication No. US20030148459A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,161
; CURRENT FILING DATE: 2002-11-21
; ERROR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: CMP-sialic acid synthetase from C. jejuni OH4384
; OTHER INFORMATION: (ORF 10a of lipooligosaccharide (LOS) biosynthesis
; OTHER INFORMATION: locus)
US-10-303-161-37

```

Query Match	7.5%	Score 169.5	DB 14	Length 221
Best Local Similarity	28.3%	Pred. No. 2e-08		
Matches	63	Conservative	35	Mismatches 106
				Indels 19
				Gaps 5

QY	47	ALILARGSGKGIPLKNIKHLGVLGVLGWLRAALDSCAEQVWVSTDHDIENVAKQFGA	105
DB	4	ALIPARGSGKGIKKNKLVLLNNKPLIYYTTKAALNTKS-SKVVVSSDSDELLNTAKSQNV	63
QY	107	QVHRRSEVSKDSTSDLAIEFYANYXNDIVGNIQATXCLXHTPTDLOKVAEMIREEGY	166
DB	64	DILKRPISLAQNTTSDKVLHLHALFKYKVEDVVFLQTSPLTRNIHIDEAFNLKNSNA	123
QY	167	DSXFSVVRHQFRWSEIQKV-----REVTEPLNLFAKEP--RRQDWGDELVENGSFY	218
DB	124	NALISV-----SECDWKILKAFVCFVGDLAGICNDIEPFWPKLPLKTYMSNGATY	175
QY	219	F--AKRHLEMGYLQGGKWHTTKCELEHSVDIDVDIDWPIAEQ	259
DB	176	ILKIKGFLLNPSFLQSKTHFLWDE-SSSDICLEDLKKAC	217

RESULT 15

US-10-303-218-37

```

; Sequence 37, Application US/10303118
; Publication No. US20030157655A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10303,118
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: CMP-sialic acid synthetase from C. jejuni CH4384
; OTHER INFORMATION: (ORF 10a of lipooligosaccharide (LOS) biosynthesis
; OTHER INFORMATION: locus)
; US-10-303-118-37

Query Match          7.5%; Score 169.5; DB 14; Length 221;
Best Local Similarity 28.3%; Pred No. 2e-08;
Matches 63; Conservative 33; Mismatches 106; Indels 19; Gaps 5

QY 47 ALIARGSGSGIPLKNIKHLAEGVLIQWLVRAALDSGAFQSVQWSTDHDEIENVAKQFCA 106
Db 4 AIIIPARGSGSGIKNNKVLNKNKPLIYVYTIKAALNTKSIKSVVSSDSDEILNLYAKSNV 63
QY 107 QVRRRSEVSKDSTSLDAIIIEFLNYKNEKDIVNGIONATSIKCLPSTDLOKVAEMTBEGY 166
Db 64 DILKRPISLAQDNNTSDKVLHLHALFKYKDYEDVVFLQPTSLRPTNIHIDEAFNLYKNSNA 123
QY 167 DSXFPVRRHQFWSIEQKGV-----REVTEPLNLPKAPRP--RRQWDGDELYENGSGFY 218
Db 124 NALLISV-----SECDNKLKAFVNCNEVGLAGI CNDEYFPMPRQKLPXTYNSNGALY 175
QY 219 F--AKRHLEPMGYLQGGKWHKTKCELEHSDVIDDVIDWPIAEQ 259
Db 176 ILKIKSFIANPFSLOSKTKEHFLMD--SSSLDIDCDLEDLKKRAEO 217

```

Search completed: May 6, 2004, 09:25:05
Job time : 47.9189 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:34 ; Search time 4393.16 Seconds
(without alignments)
10679.608 Million cell updates/sec

Title: US-09-930-440B-5

Perfect score: 1080

Sequence: 1 atgcgcgtggagc:ggagct.....gcaaaaaaatcaagctcttaa 1080

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_ov.*

5: gb_ov.*

6: gb_ov.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_or.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_p_n.*

35: em_htg_red.*

36: em_htg_nam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1080	100.0	1080	6	BD270235	BD270235 Human gly
2	1080	100.0	1080	9	AF257466	AF257466 Homo sapi
3	1076.8	99.7	1181	3	BC019315	BC019315 Homo sapi
4	1076.8	99.7	1229	9	BC000008	BC000008 Homo sapi
5	1076.8	99.7	1233	6	AR339544	AR339544 Sequence
6	1075.2	99.6	1170	6	AX877224	AX877224 Sequence
7	1075.2	99.6	1170	6	BD156549	BD156549 Primer fo
8	1075.2	99.6	1170	6	AK001659	AK001659 Homo sapi
9	883.2	81.8	1247	10	BC003307	BC003307 Mus muscu
10	883.2	81.8	1916	10	BC057977	BC057977 Mus muscu
11	881.6	81.6	1174	10	AB041263	AB041263 Mus muscu
12	732.4	67.8	3026	9	AK091184	AK091184 Homo sapi
13	710.2	65.8	813	6	AX868777	AX868777 Sequence
14	710.2	65.8	813	6	BD148839	BD148839 Primer fo
15	483.8	44.8	514	6	AX873806	AX873806 Sequence
16	483.8	44.8	514	6	BD153868	BD153868 Primer fo
17	456.6	42.3	2373	2	AC097420	AC097420 Rattus no
18	418	38.7	464	6	AX408867	AX408867 Sequence
19	386.8	35.8	553	6	AX557002	AX557002 Sequence
20	338.6	31.4	2289	2	AC106079	AC106079 Rattus no
21	338.6	31.4	2292	2	AC094185	AC094185 Rattus no
22	280.6	26.0	404	6	AX201885	AX201885 Sequence
23	269.8	25.0	2044	9	AK126144	AK126144 Homo sapi
24	269.8	25.0	2013	9	AL137073	AL137073 Human DNA
25	241.6	22.4	2278	10	AL683884	AL683884 Mouse DNA
26	241.4	22.4	2441	2	AC097073	AC097073 Rattus no
27	221	20.5	1194	3	AF397531	AF397531 Drosophil
28	210.4	19.5	1999	9	AK123739	AK123739 Homo sapi
29	199.6	18.5	1975	2	AC137022	AC137022 Rattus no
30	195	18.1	1823	2	AC122081	AC122081 Rattus no
31	195	18.1	1918	2	AC112827	AC112827 Rattus no
32	195	18.1	2664	2	AC120718	AC120718 Rattus no
33	145.6	13.5	2921	1	SC093912	SC093912 Streptomy
34	144.4	13.4	1123	1	AB012850	AB012850 Chlorobiu
35	135	12.5	3023	1	AF005034	AF005034 Streptomy
36	126.4	11.7	2980	1	BX321861	BX321861 Nitrosomo
37	125	11.6	3496	2	AB055099	AB055099 Synechoco
38	119.8	11.1	1526	3	AB055099	AB055099 Drosophil
39	119.8	11.1	2290	3	AB055101	AB055101 Drosophil
40	119.8	11.1	2307	3	AB055100	AB055100 Drosophil
41	119.8	11.1	2344	3	AB055098	AB055098 Drosophil
42	119.8	11.1	6120	2	AC017132	AC017132 Drosophil
43	119.8	11.1	1616	2	AC007594	AC007594 Drosophil
44	119.8	11.1	2256	3	AB003695	AB003695 Drosophil
45	111.8	10.4	8689	1	RCU57682	US7682 Rhodobacter

ALIGNMENTS

RESULT 1
BD270235
LOCUS BD270235 1080 bp DNA linear PAT 17-JUL-2003
DEFINITION Human glycosylation enzymes.
ACCESSION BD270235
VERSION BD270235.1 GI:33080003
KEYWORDS JP 2002537796-A/2.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1080)
AUTHORS Coleman, T.A. and Betenbaugh, M.J.
TITLE Human glycosylation enzymes
JOURNAL Patent: JP 2002537796-A 2 12-NOV-2002;

[illegible]

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E.B. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 9 Row: d Column: 7
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 12056472.

FEATURES

source

```

1..1181
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:4339 IMAGE:2821389"
/tissue_type="Lung, small cell carcinoma"
/clone_id="NH1_MGC_7"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
1..1181
/gene="NANS"
/notes="synonym: SAS"
/db_xref="LocusID:54187"
/db_xref="MIM:605202"
61..1140
/codon_start=1
/product="N-acetylneuraminic acid phosphate synthase"
/protein_id="AAH19315.1"
/db_xref="GI:17939512"
/db_xref="LocusID:54187"
/translation="MPLELELCPRVWGQHPFIIAETIGNHQGLDVAKMRIMAK
EGADCAKFQSELEKFNKALDRPYTSKHSWKTYGEHKLHFSHDYRELQRYA
SEVITFFASGDMVAFLEHNPFFKVGSDNTNFFLYLEKTAQKGRPMVISGQM
SMTWKQVQIVKLPNMPFQLOCTSAVPLQEDVNLRLVISYQKLFPDIPIGYSGE
TGALSAVALGALSERHITLDTKWTGSDHSALEPCELAELVRSVPLVERALGSP
TKQLLPCMACNEKLGKSVAVKVPBETILLMDLTVKVGEPKGPEDIFNLVKK
VLVTVEEDDTIMEELVDNHGKIKS"
82..1095
/notes="SpSE; Region: Sialic acid synthase [Cell envelope
biogenesis, outer membrane]"
/db_xref="CDD:COG2089"

```

misc_feature

ORIGIN

```

Query Match: 99.7%; Score 1076.8; DB 9; Length 1181;
Best Local Similarity 99.8%; Pred. No. 1.4e-255;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCGCTGGAGCTGGAGTGTGTCCTGGGCGCTGGTGGGGGCGCAACACCGCTGCTTC 60
DB 1 ATGCGCTGGAGCTGGAGTGTGTCCTGGGCGCTGGTGGGGGCGCAACACCGCTGCTTC 120
QY 61 ATCATTGGCGAGATCGGCAGAACCAACCAAGCGGCGACCTGGAGCTAGCCAAAGCGCATGATC 120
DB 121 ATCATTGGCGAGATCGGCAGAACCAACCAAGCGGCGACCTGGAGCTAGCCAAAGCGCATGATC 180
QY 121 CGCATGGCCAGAGTGTGGGCTGATTGTGCCAAGTTCAGAAAGGTGAGTGAATTC 180
DB 181 CGCATGGCCAGAGTGTGGGCTGATTGTGCCAAGTTCAGAAAGGTGAGTGAATTC 240
QY 181 AASTTTATCGAAGACCTTGGAGAGGCGATACACTCGAAGCTATCTCTGGGGGAGAGC 240
DB 241 AGTTTATCGAAGACCTTGGAGAGGCGATACACTCGAAGCTATCTCTGGGGGAGAGC 300
QY 241 TACGGGGAGCAAAACGACATCTGAGTTCACCCATGACCATGACAGGAGGTGCGAGAGG 300
DB 301 TACGGGGAGCAAAACGACATCTGAGTTCACCCATGACCATGACAGGAGGTGCGAGAGG 360
QY 301 TACGGGGAGGAGTGGAGTCTTCTTCTACTGCTCTGGCATGGATGAGATGGCAGTTGAA 360
DB 361 TACGGGGAGGAGTGGAGTCTTCTTCTACTGCTCTGGCATGGATGAGATGGCAGTTGAA 420
QY 361 TTCTGTGATGACTCAATGCTTCAATTTCAAGTTGGATCTGAGAGCACTAATTAATTT 420
DB 421 TTCTGTGATGACTCAATGCTTCAATTTCAAGTTGGATCTGAGAGCACTAATTAATTT 480
QY 421 CTTTATCTGGAAAGACAGCAAAAGGTCGCCCAATGGTGTATCTCCAGTGGGATCGAG 480

```

```

DB 481 CTTTATCTGGAAAGACAGCAAAAGGTCGCCCAATGGTGTATCTCCAGTGGGATGCAG 540
QY 481 TCAATGGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 540
DB 541 TCAATGGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 600
QY 541 TTCTTGCAGTGTACAGCGGCATACCCGCTCCAGCCTGAGGAGTCAACCTGGGGTCATC 600
DB 601 TTCTTGCAGTGTACAGCGGCATACCCGCTCCAGCCTGAGGAGTCAACCTGGGGTCATC 660
QY 601 TCGGATATCAGAGTCTTTTCTCTGACATTCCCATAGGGTATTCTTGGGATGAACAGGC 660
DB 661 TCGGATATCAGAGTCTTTTCTCTGACATTCCCATAGGGTATTCTTGGGATGAACAGGC 720
QY 661 ATAGCGATATCTGTGGCCGACGTGGCTCTGGGGGCCAAGGTGTGTGAACCTCACATTA 720
DB 721 ATAGCGATATCTGTGGCCGACGTGGCTCTGTGGGGGCCAAGGTGTGTGAACCTCACATTA 780
QY 721 TTGGACACACCTGGAGGGGATGACCACTGGGCTCTGCTGGAGCTTGGAGAGCTGGAGGC 780
DB 781 TTGGACACACCTGGAGGGGATGACCACTGGGCTCTGCTGGAGCTTGGAGAGCTGGAGGC 840
QY 781 GAGCTGGTGGCTGAGTGTGCTGTGTGGAGCGTGGCTGGGCTCCCAACCAAGCAGCTG 840
DB 841 GAGCTGGTGGCTGAGTGTGCTGTGTGGAGCGTGGCTGGGCTCCCAACCAAGCAGCTG 900
QY 841 CTGCGCTGTGAGATGGCTTGCATGAGAGCTGGGCAAGTCTGTGTGGGCTCAAAAGTGAA 900
DB 901 CTGCGCTGTGAGATGGCTTGCATGAGAGCTGGGCAAGTCTGTGTGGGCTCAAAAGTGAA 960
QY 901 ATTCCGGAGGACCACTTCTACATGAGTGTGCTACCGTGAAGTGGTGGAGCCAA 960
DB 961 ATTCCGGAGGACCACTTCTACATGAGTGTGCTACCGTGAAGTGGTGGAGCCAA 1020
QY 961 GCCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGAGTCTGTGTCTGCTGCTGAA 1020
DB 1021 GCCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGAGTCTGTGTCTGCTGCTGAA 1080
QY 1021 GAGGTGACACCATGATGAGAGATTTGATGATATCATGTCGCAAAATACTGCTTTAA 1080
DB 1081 GAGGTGACACCATGATGAGAGATTTGATGATATCATGTCGCAAAATACTGCTTTAA 1140

RESULT 4
BC000008 1229 bp mRNA linear PRI 04-OCT-2003
Homo sapiens N-acetylneuraminic acid synthase (sialic acid
synthase), mRNA (cdna clone MGC:934 IMAGE:3505324), complete cds.
ACCESSION BC000008
VERSION BC000008.2 GI:33875092
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1229)
Straussberg,R.B., Feingold,B.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L.,
Scheet,R.B., Brownstein,M.J., Usdin,R.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raja,S., Loquellano,N.A., Peters,G.J.,
Abramson,K.J., Mullah,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Farey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,S.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,

```

TITLE Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 1229)
Strausberg, R.

TITLE Direct Submission
Submitted (03-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT On Aug 19, 2003 this sequence version replaced gi:126525238.

Contact: MGC help desk
 Email: cqapbs-remail.nih.gov

Tissue Procurement: ATCC
cDNA Library Preparation

CDNA Library Arrayed by: The I.M.A.G.E. Consortium
Library Preparation: Novus Laboratory
DNA Sequencing by: Institute for Systems Biology
(LNL)

http://www.systemsbiology.org
contact: amadan@systemsbiology.org

CONTACT: amuradha@stanford.edu, 650.725.1829
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

C-clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.linl.gov>
 Series: IRAL P.M.E. 7 Row: f Column: 1
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mpv2017_12055472

FEATURES

source

1. 1229
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:934 IMAGE:3505324"
 /tissue_type="Placenta, Choriocarcinoma"
 /clone_lib="NIH_MGC_21"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 1. 1229
 /gene="NANKS"
 /note="synonym: SAS"
 /db_xref="LocusID:54187"
 /db_xref="XIN:605202"
 79. 1158
 /codon_start=1
 /product="N-acetylneuraminic acid phosphate synthase"
 /protein_id="AAH00008.1"
 /db_xref="GI:12652539"
 /db_xref="LocusID:54187"
 /translation="SLELELCPCRWGGQHPFCITAEIGNHQGLDVAKRMIRMAX
 ECGACAKQKSLSEFKFKALKALEPVTSKHSGKTLQYGBKHLEFSDQYRELQRYA
 ZEYGIPTFASGDMDEAVFELHNLNVFPFKVSGSDTNFYLEKTIARAGKAPWVLISSGQ
 SMDTKQVYVAIPLKPNFCFLQCTSAIPLQPDVNLRISEYQKLFDPDIPITYSGHE
 TGIATQVAALGAKLERHITLDKTKWGSCHSASLGPGLAEIIVRSALVERALGSP
 TKQLLPCGNACNEKLGKSVVAKVKIPGTLITLMDLTVKVGEPKGYPPEDIPNLVCKK
 VLVTVREDPTIMEFIVDNHGKTKTS"

ORIGIN

```
Query Watch 99.7%; Score 1076.8; DB 9; Length 1229;
Best Local Similarity 99.8%; Pred. No. 1.4e-255;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 ATGCCGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGGAACAACCCGTGCTTC 60

Db 79 ATGCCGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGGAACAACCCGTGCTTC 138

61	QY	ATCATTTCCCGAGATCGGCCAGAACACACAGGCGACCTGGAGCTAGCCAGACGCAATGATC	120
139	DB	ATCATTTCCCGAGATCGGCCAGAACACACAGGCGACCTGGATGTAGCCACGACGATGATC	198
121	QY	CGCATGCCCAAGAGATGTGGGGCTGATTGTGCCAAGTTCCAGAGAGAGTGAGCTAGAAATTC	180
199	DB	CGCATGCCCAAGAGATGTGGGGCTGATTGTGCCAAGTTCCAGAGAGAGTGAGCTAGAAATTC	258
181	QY	AAGTTTAAATCGGAAAGCCTTTGGAGAGGCCMTACACCTCGAAGCAATTCCTGGGGGAAAGACG	240
259	DB	AAGTTTAAATCGGAAAGCCTTTGGAGAGGCCMTACACCTCGAAGCAATTCCTGGGGGAAAGACG	318
241	QY	TACGGGAGGACAAAAGACATCTGGAGTTTCGCCATGACCCAGTACAGGAGCTGCGAGGG	300
319	DB	TACGGGAGGACAAAAGACATCTGGAGTTTCGCCATGACCCAGTACAGGAGCTGCGAGGG	378
301	QY	TACGCCGAGAGGTTGGGATCTTCTTCACTGCGCTCTGGCATGGATGAGATGGCAGTTGAA	360
379	DB	TACGCCGAGAGGTTGGGATCTTCTTCACTGCGCTCTGGCATGGATGAGATGGCAGTTGAA	438
361	QY	TTCTCTGCATGAACTGAAATGTTCCATTTTCAAGTTGGATCTCGAGACACTAAATAATTTT	420
439	DB	TTCTCTGCATGAACTGAAATGTTCCATTTTCAAGTTGGATCTCGAGACACTAAATAATTTT	498
421	QY	CTTTATCTGAAAGAAGCAGCCAAAAGAGTTCGCCCAATGGTGATCTCCAGTGGGATCGAG	480
499	DB	CTTTATCTGAAAGAAGCAGCCAAAAGAGTTCGCCCAATGGTGATCTCCAGTGGGATCGAG	558
481	QY	TCAATGACACCCATGAGCAAGTTTATCAGATCTGAAAGCCCTTCAACCCCAACTCTTCGC	540
559	DB	TCAATGACACCCATGAGCAAGTTTATCAGATCTGAAAGCCCTTCAACCCCAACTCTTCGC	618
541	QY	TTCTTGAGTGTACGAGCGCATACCGCTCCAGCTCGAGAGCTGATCAACTCGCGGGTCATC	600
619	DB	TTCTTGAGTGTACGAGCGCATACCGCTCCAGCTCGAGAGCTGATCAACTCGCGGGTCATC	678
601	QY	TCGGATATCAGAAAGCTCTTTCTCTGCATCTCCATAGGTTATCTCGGCGATGAAACAGGC	660
679	DB	TCGGATATCAGAAAGCTCTTTCTCTGCATCTCCATAGGTTATCTCGGCGATGAAACAGGC	738
661	QY	ATAGCGATATCTGTGCCCGCAGTGGCTCTCGGGGCCAAGGTGTGGAACTGACATAACT	720
739	DB	ATAGCGATATCTGTGCCCGCAGTGGCTCTCGGGGCCAAGGTGTGGAACTGACATAACT	798
721	QY	TTGGAACAAGACTGGAAGGGAGTGCACCTCGGCCCTCGCTGAGAGCTGAGAGACTGGCC	780
799	DB	TTGGAACAAGACTGGAAGGGAGTGCACCTCGGCCCTCGCTGAGAGCTGAGAGACTGGCC	858
781	QY	GAGCTGGTGGGTCACTGGCTCTGTGGAGCGTTCGCTGGGCTCCCCAAACAGCAGCTG	840
859	DB	GAGCTGGTGGGTCACTGGCTCTGTGGAGCGTTCGCTGGGCTCCCCAAACAGCAGCTG	918
841	QY	CTGCCCTGTGAGATGCCCTGCMAATGAGAAGCTGGGCAAGTCTCTGTGGTGGCCAAAGTGA	900
919	DB	CTGCCCTGTGAGATGCCCTGCMAATGAGAAGCTGGGCAAGTCTCTGTGGTGGCCAAAGTGA	978
901	QY	ATTCCGGAAGGACCACTTCTTACAAATGGACATGCTCACCGTGAAGTGGCTGAGCCCAA	960
979	DB	ATTCCGGAAGGACCACTTCTTACAAATGGACATGCTCACCGTGAAGTGGCTGAGCCCAA	1038
961	QY	GCTATCTCTCTGAAGACATCTTTTAATCTAGTGGGCAAGAAGTCTCTGGTCACTGTGAA	1020
1039	DB	GCTATCTCTCTGAAGACATCTTTTAATCTAGTGGGCAAGAAGTCTCTGGTCACTGTGAA	1098
1021	QY	GAGGATGACCACTCATGAAGAATTTGATAGATAATCATGGCAAAAAAATCAAGTCTTAA	1080
1099	DB	GAGGATGACCACTCATGAAGAATTTGATAGATAATCATGGCAAAAAAATCAAGTCTTAA	1158

RESULT 5
AR339544
LOCUS
DEFINITION

```
ACCESSION AR339544
VERSION AR339544.1 GI:33726401
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1230)
AUTHORS Tang, Y.T., Zhou, P. and Dmanac, R.T.
TITLE Nucleic acids and polypeptides
JOURNAL Patent: US 659662-A 1235 27-MAY-2003;
FEATURES Location/Qualifiers
source 1..1230
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 99.7%; Score 1076.8; DB 6; Length 1230;
Best Local Similarity 99.8%; Pred. No. 1.4e-255;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCCGTGGAGCTGGAGCTGTGTCGCCGGCGCTGGTGGCGGGCAACACCCGTGCTTC 60
Db 119 ATGCCGTGGAGCTGGAGCTGTGTCGCCGGCGCTGGTGGCGGGCAACACCCGTGCTTC 178
Qy 61 ATCAATCCGAGATCGCCCAACACCCAGCGGACCTGGACGTAGCCCAAGCGCATGATC 120
Db 179 ATCAATCCGAGATCGCCCAACACCCAGCGGACCTGGACGTAGCCCAAGCGCATGATC 238
Qy 121 CGCATGCCAAGAGTGTGGGGCTGATGTGCCAAGTTCAGAGAGAGTGAGCTAGATTC 180
Db 239 CGCATGCCAAGAGTGTGGGGCTGATGTGCCAAGTTCAGAGAGAGTGAGCTAGATTC 298
Qy 181 AAGTTTAAATCGGAAGCTTGGAGAGCCATACACCTCGAAGCTTCTTGGGGGAAGACG 240
Db 299 AAGTTTAAATCGGAAGCTTGGAGAGCCATACACCTCGAAGCTTCTTGGGGGAAGACG 358
Qy 241 TACGGGAGCAACACGACATCTGGAGTTCAGCCATGACAGTACAGGAGCTGCAGAGG 300
Db 359 TACGGGAGCAACACGACATCTGGAGTTCAGCCATGACAGTACAGGAGCTGCAGAGG 418
Qy 301 TACGCCAGGAGGTGGGATTTCTTCACTGCCCTTGGCATGGATGGATGGCGATGAA 360
Db 419 TACGCCAGGAGGTGGGATTTCTTCACTGCCCTTGGCATGGATGGATGGCGATGAA 478
Qy 361 TCCCTGATGAACCTGAATGTTCCATTTTCAAGTTCAGATTCGATCGGACACTAAATTT 420
Db 479 TCCCTGATGAACCTGAATGTTCCATTTTCAAGTTCAGATTCGATCGGACACTAAATTT 538
Qy 421 CCTTATCTGAAAAGACAGCCAAAAGGTCGCCAATGGTGTATCTCCAGTGGGATGAG 480
Db 539 CCTTATCTGAAAAGACAGCCAAAAGGTCGCCAATGGTGTATCTCCAGTGGGATGAG 598
Qy 481 TCAATGACACCATGAGCAAGTTTATCAGATCTGTAAGCCCTCAACCCCAACTTCGC 540
Db 599 TCAATGACACCATGAGCAAGTTTATCAGATCTGTAAGCCCTCAACCCCAACTTCGC 658
Qy 541 TCTCTGAGTGTACAGCGATACCCGCTCCAGCTGAGGAGCTCAACCTTGGGTCATC 600
Db 659 TCTCTGAGTGTACAGCGATACCCGCTCCAGCTGAGGAGCTCAACCTTGGGTCATC 718
Qy 601 TCGGAATATCAGAGCTCTTCTGACATTCCTGATGATTCCTGGGATGAAACAGGC 660
Db 719 TCGGAATATCAGAGCTCTTCTGACATTCCTGATGATTCCTGGGATGAAACAGGC 778
Qy 661 ATAGCGATATCTGTGGCGAGTGGCTCTGGGGGCAAGGTTGTGGAACGTCAATACT 720
Db 779 ATAGCGATATCTGTGGCGAGTGGCTCTGGGGGCAAGGTTGTGGAACGTCAATACT 838
Qy 721 TTGACAGACCTCGAAGGGAGTGACCACTCGGCTCGCTGGAGCTCGGAGAACTGCC 780
Db 839 TTGACAGACCTCGAAGGGAGTGACCACTCGGCTCGCTGGAGCTCGGAGAACTGCC 898
Qy 781 GAGCTGGTGCAGTGCAGTCTTGTGGAGGCTGCCCTGGGCTCCCAACCAAGCAGCTG 840
```

```

999 GAGCTGGTGCAGTGCAGTCTTGTGGAGGCTGCCCTGGGCTCCCAACCAAGCAGCTG 958
841 CTGCCCTGTGAGATGGCTCAATGAGAGAGTGGGCAAGTCTGGTGGCCAAAGTAAA 900
959 CTGCCCTGTGAGATGGCTCAATGAGAGAGTGGGCAAGTCTGGTGGCCAAAGTAAA 1018
901 ATTCCGGAAGCAGCATCTTACCAATGACATGCTCCCGTGAAGGTGGGTGGGCGCCAAA 960
1019 ATTCCGGAAGCAGCATCTTACCAATGACATGCTCCCGTGAAGGTGGGTGGGCGCCAAA 1078
961 GCCTATCTCTCTGAAGCATCTTTAACTAGTGGGCAAGAGGTCTCTGCTCACTGTTCAA 1020
1079 GCCTATCTCTCTGAAGCATCTTTAACTAGTGGGCAAGAGGTCTCTGCTCACTGTTCAA 1138
1021 GAGATGACACCATCATCATGGAAGAAATGTGTAGATATCATGCAAAAATAATCAAGTCTTAA 1080
1139 GAGATGACACCATCATCATGGAAGAAATGTGTAGATATCATGCAAAAATAATCAAGTCTTAA 1198

RESULT 6
AX877224 1170 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 12129 from Patent EP1074617.
ACCESSION AX877224
VERSION AX877224.1 GI:40031960
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Negai, K. and Otsuki, T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 12129 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES Location/Qualifiers
source 1..1170
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
71..1150
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE89627.1"
/db_xref="GI:40031961"
/translacion="VPLELELCPORWVGQHPFIIABIGNHQDLDVAKRMIRMAX
ECGADCAKFKSELEFPNFKALDRPYTSKHSWGKTYGHEKRELEFSDHQVREIQRYA
BEVGIFPTAGMDMAVEFLHNLNPPFKVSGSDTNPPYLEKTAKKRPMWISGGMQ
SMOTMKOVYQIVKPLNPFCLQCTSAVPIQPEDVNLNVISEYOKLPDIPIGYSGHE
TGIAISVAVALGTVLERTHTLDTKWKSGDSHSLSPGLASLWRSVRLVERALGSP
TKQLLPCEMACNEKLKGSVWAKVXIPETGTLITMDMLTVKVGEPKGYPPEDIFNLVKK
VLVTVEEDDTIMEELVDNHHKKIXS"

ORIGIN
Query Match 99.6%; Score 1075.2; DB 6; Length 1170;
Best Local Similarity 99.7%; Pred. No. 3.4e-255;
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGCCGTGGAGCTGGAGCTGTGTCGCCGGCGCTGGTGGCGGGCAACACCCGTGCTTC 60
Db 71 ATGCCGTGGAGCTGGAGCTGTGTCGCCGGCGCTGGTGGCGGGCAACACCCGTGCTTC 130
Qy 61 ATCATTCGCGAGATCGCCCAAGAACCAACAGGCGACCTGGACCTAGCCAAAGCATGATC 120
Db 131 ATCATTCGCGAGATCGCCCAAGAACCAACAGGCGACCTGGACCTAGCCAAAGCATGATC 190
Qy 121 GCGATGCCAAGAGGTGTGGGGCTGATGTGCCAAGTTCAGAGAGAGTGAGCTAGATTC 180
Db 191 GCGATGCCAAGAGGTGTGGGGCTGATGTGCCAAGTTCAGAGAGAGTGAGCTAGATTC 250
Qy 181 AAGTTTAAATCGGAAGCTTGGAGAGGCAATACACCTCGAAGCAATCTCTGGGGGAAGACG 240
```

251	Db		AAAGTTTAATCGGAAAGCCTTGGACAGGCCATACACCTCGAAGCAATTCCTGGGGGAAGACG	310
241	QY		TACGGGAGACACAAACGACATCTGGAGTTCAGCCATGACACAGTACAGGAGCTGCAGAGG	300
311	Db		TACGGGAGACACAAACGACATCTGGAGTTCAGCCATGACACAGTACAGGAGCTGCAGAGG	370
301	QY		TACGCCGAGAGGTTGGATCTTCTTCACTGCCTCTGGCATGGATGAGATGGCAGTTGAA	360
371	Db		TACGCCGAGAGGTTGGATCTTCTTCACTGCCTCTGGCATGGATGAGATGGCAGTTGAA	430
361	QY		TTCTCTGCATGAACTGAATGTCCATTTTCAAGTTGATCTGAGACACTAATAATTTT	420
431	Db		TTCTCTGCATGAACTGAATGTCCATTTTCAAGTTGATCTGAGACACTAATAATTTT	490
421	QY		CCTTATCTGAAAGACAGCCAAAGGTCGCCAATGGTGATCTCCAGTGGGATCGAG	480
491	Db		CCTTATCTGAAAGACAGCCAAAGGTCGCCAATGGTGATCTCCAGTGGGATCGAG	550
481	QY		TCAATGACACCAATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAAACCCAACTTCGC	540
551	Db		TCAATGACACCAATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAAACCCAACTTCGC	610
541	QY		TTCTTGCAGTGTACCGACGATACCCGCTCCAGCTCAGGACGTCAACCTCGCGGTCATC	600
611	Db		TTCTTGCAGTGTACCGACGATACCCGCTCCAGCTCAGGACGTCAACCTCGCGGTCATC	670
601	QY		TCGGAATATCAGAAAGCTCTTTTCTGTGACATTCCTAGAGTATCTGGGCATGAAACAGGC	660
671	Db		TCGGAATATCAGAAAGCTCTTTTCTGTGACATTCCTAGAGTATCTGGGCATGAAACAGGC	730
661	QY		ATAGCGATATCTGTGGCCGAGTGGCTCTGGGGCCAAAGTTGTGGAACTCACATAACT	720
731	Db		ATAGCGATATCTGTGGCCGAGTGGCTCTGGGGCCAAAGTTGTGGAACTCACATAACT	790
721	QY		TTGGACAGACCTGGAGGGAGTGAACCATCGGCTTCGCTGGAGCTGGAGACTGGCC	780
791	Db		TTGGACAGACCTGGAGGGAGTGAACCATCGGCTTCGCTGGAGCTGGAGACTGGCC	850
781	QY		GAGCTGGTGGGTCAAGTCGCTGTGTGGAGCGTCCCTGGGCTCCCAACCAAGCAGCTG	840
851	Db		GAGCTGGTGGGTCAAGTCGCTGTGTGGAGCGTCCCTGGGCTCCCAACCAAGCAGCTG	910
841	QY		CTGCCCTGTGAGATGGCCTCCAATGAGAACTGGGCAAGTCTGTGTGGGCCAAAGTGAAA	900
911	Db		CTGCCCTGTGAGATGGCCTCCAATGAGAACTGGGCAAGTCTGTGTGGGCCAAAGTGAAA	970
901	QY		ATTCCGGAGGACCAATCTTAACTAATGGACATGCTACCGTGAAGTGGGTGAGCCCAA	960
971	Db		ATTCCGGAGGACCAATCTTAACTAATGGACATGCTACCGTGAAGTGGGTGAGCCCAA	1030
961	QY		GCCTATCCTCCTGAAGACATCTTTAATCTAGTGGGCAAGAGGTCTCTGGTCACTGTTGAA	1020
1031	Db		GCCTATCCTCCTGAAGACATCTTTAATCTAGTGGGCAAGAGGTCTCTGGTCACTGTTGAA	1090
1021	QY		GAGGATGACACCAATCATGGAAGAAATTTGGTATATCATGCGCAAAAAATCAAGTCTTAA	1080
1091	Db		GAGGATGACACCAATCATGGAAGAAATTTGGTATATCATGCGCAAAAAATCAAGTCTTAA	1150

Ishii,S., Sugiyama,T., Wakamatsu,A., Negai,K. and Otsuki,T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 11392 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/11392
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU
PI KEIICHI NAGAI,TETSUJI OTSUKI
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N3/21,C12N5/PC
10,
PC C12P21/02,C1201/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof PH Key
Location/Qualifiers
FT CDS {71}..(1147).
Location/Qualifiers
1..1170
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 99.6%; Score 1075.2; DB 6; Length 1170;
Best Local Similarity 99.7%; Pred. No. 3.4e-255;
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0

Qy 1 ATGCCGTGGAGCTGGAGCTGTGTCCCGGGCGCTGGGTGGCGCGGCAACACCCGCGCTTC 60
Db 71 ATGCCGTGGAGCTGGAGCTGTGTCCCGGGCGCTGGGTGGCGCGGCAACACCCGCGCTTC 130

Qy 61 ATCAATGCCAGATCGGCCAGAACCAACAGCGGCACTGGACGTAGCCAAAGGCGATGATC 120
Db 131 ATCAATGCCAGATCGGCCAGAACCAACAGCGGCACTGGACGTAGCCAAAGGCGATGATC 190

Qy 121 CGCATGCCAAGAGTGTGGGGCTGATTGTGCCAAGTTCACAGAAGAGTGAGTAGAATTC 180
Db 191 CGCATGCCAAGAGTGTGGGGCTGATTGTGCCAAGTTCACAGAAGAGTGAGTAGAATTC 250

Qy 181 AGTTTAAATCGAAAGCCTTGGAGAGCCATACACCTCGAAGCATTCCTGGGGGAAGACG 240
Db 251 AGTTTAAATCGAAAGCCTTGGAGAGCCATACACCTCGAAGCATTCCTGGGGGAAGACG 310

Qy 241 TAGGGGAGACAAACGACATCTGGAGTTCAGCCATGACCCAGTACAGGAGCTCGAGAGG 300
Db 311 TAGGGGAGACAAACGACATCTGGAGTTCAGCCATGACCCAGTACAGGAGCTCGAGAGG 370

Qy 301 TAGCCGAGAGGTTGGGATCTCTTCACTGCCCTCTGGCATGGATGAGATGCGCAGTTGAA 360
Db 371 TAGCCGAGAGGTTGGGATCTCTTCACTGCCCTCTGGCATGGATGAGATGCGCAGTTGAA 430

Qy 361 TTCTGTGATGAACCTGAATGTTCATTTTCAAAGTTGGATCTGGAGACACTAATAATTTT 420
Db 431 TTCTGTGATGAACCTGAATGTTCATTTTCAAAGTTGGATCTGGAGACACTAATAATTTT 490

Qy 421 CCTTATCTGAAAGACAGCCAAAGGTGCGCCCAATGGTGATCTCCAGTGGGATGCGAG 480
Db 491 CCTTATCTGAAAGACAGCCAAAGGTGCGCCCAATGGTGATCTCCAGTGGGATGCGAG 550

Qy 481 TCAATGACACCATGAAGCAGTTTATCAGATCTGTGAAGCCCTCAACCCCACTTCTGCG 540
Db 551 TCAATGACACCATGAAGCAGTTTATCAGATCTGTGAAGCCCTCAACCCCACTTCTGCG 610

Qy 541 TTCTTGAGTGTACCGCGCATACCCGCTCCAGCCTCGAGGACGTCACACCTCGCGGTCA 600
Db 611 TTCTTGAGTGTACCGCGCATACCCGCTCCAGCCTCGAGGACGTCACACCTCGCGGTCA 670

Qy 601 TCGAATATCAGAGCTCTTTCTCGACATTCCTCCATAGGGTATTCTGGGCGATGAACAGGC 660
Db 671 TCGAATATCAGAGCTCTTTCTCGACATTCCTCCATAGGGTATTCTGGGCGATGAACAGGC 730

```

QY 661 ATAGCGATATCTGCGCGCACTGCTCTGGGGGCAAGGTGTGGAAGCTCACATAACT 720
Db 731 ATAGCGATATCTGCGCGCACTGCTCTGGGGGCAAGGTGTGGAAGCTCACATAACT 790
QY 721 TTGGACAAGACCTGGAAGGGAGTGACCACTCGGCTCGCTGAGAGCTGGAGACTGGCC 780
Db 791 TTGGACAAGACCTGGAAGGGAGTGACCACTCGGCTCGCTGAGAGCTGGAGACTGGCC 850
QY 781 GAGCTGGTGGCTGAGTGGTCTGTGTGGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 840
Db 851 GAGCTGGTGGCTGAGTGGTCTGTGTGGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 910
QY 841 CTGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
Db 911 CTGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 970
QY 901 ATTCGGGAAGCACCATCTTACAAATGGACATGCTCACCGTGAAGTGGTGGTGGTGGTGG 960
Db 971 ATTCGGGAAGCACCATCTTACAAATGGACATGCTCACCGTGAAGTGGTGGTGGTGGTGG 1030
QY 961 GCCTATCTCTCTGAAGACATCTTTTAACTTAGTGGGCAAGAGGTCTCTGCTCACTGTGAA 1020
Db 1031 GCCTATCTCTCTGAAGACATCTTTTAACTTAGTGGGCAAGAGGTCTCTGCTCACTGTGAA 1090
QY 1021 GAGGATGACACCATCTGGAAGATTTGGTGAATATCATGTCGCAAAATCAAGTCTTAA 1080
Db 1091 GAGGATGACACCATCTGGAAGATTTGGTGAATATCATGTCGCAAAATCAAGTCTTAA 1150

RESULT 8
AK001659
LOCUS
DEFINITION
Homo sapiens cDNA FLJ10797 fis, clone NT2RP4000657, weakly similar
to SPORE COAT POLYSACCHARIDE BIOSYNTHESIS PROTEIN SPSE.
AK001659
ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Aotsuka, S., Yoshikawa, Y.,
Matsumura, E., Ishii, S., Kawai, Y., Saito, K., Yamamoto, S.,
Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y. and Sasaki, N.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 1170)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction; 5' & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
source
1..1170
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP4000657"
/cell_line="NT2"
/cell_type="eratocarcinoma"
/clone_lib="NT2RP4"
/note="Cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)

```

```

induction."
71..1150
/codon_start=1
/feature="unlabeled protein product"
/db_xref="GI:7023453"
/translation="MPELELCPCRWVGGQHPFIIAETGQHQDLDVAKMIRMAK
EGADCAKFKQSELEFKNRKLDRPYTSKHSWGDNTNNFFYLKTKAKGPMWISSGQ
BEVGLFTASGMDENAVBFHELNVFFKVSQDNTNNFFYLKTKAKGPMWISSGQ
SMDTKQVQIVKPLNPFQCTSAVLPQEDNLRVISEYKLFPIPIGYSQHE
TNTISVAVALGTLVLRHITLDTKTKWSGDSASLEPGELAEIVSRVLRALGSP
TKQLLPBMACNEKLGKSVAKVTPBETILTMMLTVKVGPKGYPPEIDFNLVSKK
VLVTVEEDDTIMEELVDNHRKIKS"
ORIGIN
Query Match
Best Local Similarity 99.6%; Score 1075.2; DB 9; Length 1170;
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGCCGCTGGAGCTGGAGCTGTGTCCGCGCTGGTGGTGGGCGGCAACACCCGCTGCTTC 60
Db 71 ATGCCGCTGGAGCTGGAGCTGTGTCCGCGCTGGTGGTGGGCGGCAACACCCGCTGCTTC 130
QY 61 ATCATTCGCGGATCGCGCAGAACCAACAGGCGGCTTGGACGTAGCCCAAGCGCATGTC 120
Db 131 ATCATTCGCGGATCGCGCAGAACCAACAGGCGGCTTGGACGTAGCCCAAGCGCATGTC 190
QY 121 CGCATGGCAAGGAGTGTGGGCTGATTTGTGCCAAGTTCACAGAGAGTGCAGTAGAATTC 180
Db 191 CGCATGGCAAGGAGTGTGGGCTGATTTGTGCCAAGTTCACAGAGAGTGCAGTAGAATTC 250
QY 181 AAGTTTATCGGAAGCTTGGAGAGGCGTACACTCGAAGCATTCCTGGGGGAGACG 240
Db 251 AAGTTTATCGGAAGCTTGGAGAGGCGTACACTCGAAGCATTCCTGGGGGAGACG 310
QY 241 TACGGGAGACCAACAGCATCTCGAGTTCAGCCATACCACTACAGGAGCTGCGAGG 300
Db 311 TACGGGAGACCAACAGCATCTCGAGTTCAGCCATACCACTACAGGAGCTGCGAGG 370
QY 301 TACGCCGAGGAGTGGGATCTTTCTTCACTGCTCTGGCATGGATGAGATGGCATGAA 360
Db 371 TACGCCGAGGAGTGGGATCTTTCTTCACTGCTCTGGCATGGATGAGATGGCATGAA 430
QY 361 TTCTGCTGATGACTGAATGTTCCATTTTCAAGTGTGATCTGGAGACACTAATATTTT 420
Db 431 TTCTGCTGATGACTGAATGTTCCATTTTCAAGTGTGATCTGGAGACACTAATATTTT 490
QY 421 CCTTATCTGGAAGAGACAGCCCAAAAGGTGCGCCATGGTGTCTCCAGTGGGATGCAG 480
Db 491 CCTTATCTGGAAGAGACAGCCCAAAAGGTGCGCCATGGTGTCTCCAGTGGGATGCAG 550
QY 481 TCAATGGACACCTGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCG 540
Db 551 TCAATGGACACCTGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCG 610
QY 541 TTCTTGAGTGTACAGCGGATACCCGCTCCAGCTGAGGAGCTCAACCTGGGCTATC 600
Db 611 TTCTTGAGTGTACAGCGGATACCCGCTCCAGCTGAGGAGCTCAACCTGGGCTATC 670
QY 601 TCGGAATATCAGAGCTCTTTCTCTGATTCCTCAGATTCCTGAGGATTTCTGGCATGAAACAGGC 660
Db 671 TCGGAATATCAGAGCTCTTTCTCTGATTCCTCAGATTCCTGAGGATTTCTGGCATGAAACAGGC 730
QY 661 ATAGCGATATCTGCGCGCACTGCTCTGGGGGCAAGGTGTGGAAGCTCACATAACT 720
Db 731 ATAGCGATATCTGCGCGCACTGCTCTGGGGGCAAGGTGTGGAAGCTCACATAACT 790
QY 721 TTGGACAAGACCTGGAAGGGAGTGACCACTCGGCTCGCTGAGAGCTGGAGACTGGCC 780
Db 791 TTGGACAAGACCTGGAAGGGAGTGACCACTCGGCTCGCTGAGAGCTGGAGACTGGCC 850
QY 781 GAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 840

```

```

Db      851 GAGCTGGTGGCTGAGTGGTCTTGTGGAGCGTGGCTGGCTCCCAACCAAGCAGCTG 910
QY      841 CTGCCCTGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGTGCCCAAGTGAAA 900
Db      911 CTGCCCTGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGTGCCCAAGTGAAA 970
QY      901 ATTCCGAAGGACCACTTCAATGAGATGTCACCGTCAAGGTGGGTGAGCCCAA 960
Db      971 ATTCCGAAGGACCACTTCAATGAGATGTCACCGTCAAGGTGGGTGAGCCCAA 1030
QY      961 GCCTATCTCTCTGGAAGACATCTTTAATCTAGTGGGCAAGGCTCTGTCTACTGTGAA 1020
Db      1031 GCTATCTCTCTGGAAGACATCTTTAATCTAGTGGGCAAGGCTCTGTCTACTGTGAA 1090
QY      1021 GAGGATGACACCATCATGAGAGATTTGTAGATATCATGCGCAAAAATCAAGTCTTAA 1080
Db      1091 GAGGATGACACCATCATGAGAGATTTGTAGATATCATGCGCAAAAATCAAGTCTTAA 1150

RESULT 9
LOCUS   BC003307               1247 bp      mRNA      linear      ROD 03-OCT-2003
DEFINITION Mus musculus N-acetylneuraminic acid synthase [sialic acid
ACCESSION BC003307
VERSION   BC003307.1   GI:13097041
KEYWORDS  MGC.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 1247)
AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heish,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullen,L.S., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.W.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Stevchenko,Y.,
            Bouffard,G.O., Blakesley,K.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalius,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.,
            Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
            Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
            22386257
JOURNAL   12477932
PUBMED    2 (bases 1 to 1247)
AUTHORS   Strausberg,R.
TITLE     Direct Submission
JOURNAL   Submitted (20-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: ggaabs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu

```

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILNL at: <http://image.llnl.gov>
Series: IRAX plate: 6 Row: 1 Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 16716466.

FEATURES

Location/Qualifiers

Source

```

1..1247
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="129.C57BL/6J.FVB/N"
    /db_xref="taxon:10090"
    /clone="MGC:6982 IMAGE:3154822"
    /tissue_type="Mammary tumor. Brcal-/-fl; MMTV-Cre model. 1.0
    months old, gross tissue."
    /clone_lib="NCI CGAP_Mam3"
    /lab_host="DH10B"
    /note="Vector: pCMV-SPORT6"
1..1247
    /gene="Nans"
    /db_xref="LocusID:94181"
    /db_xref="GI:13097042"
    /transcript="MPLERUCPCRWVGSKHPCFLIAELIGNOHQDIDVAKRMITAK
    ECGADCAKFKSELEFKENKALRPYTKSHSWKTHGEHKEHLEFSDHQBLQSYA
    QEIGFTASMDMAVEFHLNVPFKVSGDTNPPYLEKTKAGRPWVSSGMO
    SMDTKOVYQIVKPLNPFLOCTSAIPLOPEDANLRVISEYOKLPDIPDTGSGHE
    TGAISVAVALGAKVLRTILDKTWKGSDHSASLFGELAEVLRSVRLVERALGSP
    KOLLPECDMAENKLGKSVVAKVKIPAGTTITLMDLTVKVGEKGYPPEDIFNLAKK
    VLVTIEBDTVMESVESHKCKNA"
145..1176
    /note="Sp8B; Region: Sialic acid synthase [Cell envelope
    biogenesis, outer membrane]"
    /db_xref="CDD:COG2089"

```

misc_feature

ORIGIN

```

Query Match      81.8%; Score 883.2; DB 10; Length 1247;
Best Local Similarity 88.6%; Pred. NC. 1.2e-207;
Matches 957; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY      1 ATCCCGCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGCGCAACACCCGTGCTTC 60
Db      124 ATCCCGCTGGAACCTGGAGCTGTGTCCCGGCGCTGGGTGGGTGGTGAAGACACCCGTGCTTC 183

QY      61 ATCAATTCGCGAGATCGCCGAGAACCCACGAGGCGAAGCTGGACCTAGCCAAAGCGCATGATC 120
Db      184 ATCAATTCGCGAGATCGCCGAGAACCCACGAGGAGACATAGATGTGGCCAAAGCGCATGATC 243

QY      121 CGCATGCCCAAGAGTGTGGGCTGATTGTGCCAAGTTCAGAGAGTCCAGAGAGTAGTAGTAATTC 180
Db      244 CGCATGCCCAAGAGTGTGGGCTGATTGTGCCAAGTTCAGAGAGTCCAGAGAGTAGTAGTAATTC 303

QY      181 AAGTTTAATCGGAAAGCCCTTGGAGAGCCCATACCTCGAAGCATTCCTGGGGGGAAGACG 240
Db      304 AAGTTTAATCGGAAAGCCCTTGGAGAGCCCATATCTTCGAGAGATTCATCGGGGGAAGACG 363

QY      241 TACGGGAGACAAACGACATCTGGAGTTCAGCCATGACCACTACAGAGAGTCCAGAGG 300
Db      364 TATGGGAGACAAACGAGCGCATCTGGAAATTCAGCCACGACCAAGTACAGAGAGTCCAGAGG 423

QY      301 TACGCCGAGAGGTTGGGATCTTCTCACTGGCTCTGGCATGGATGAGATGGCAGTTGAA 363
Db      424 TATGGCAGAGAGATCGGCATCTTCTCACTGGCATGGACGAGATGCGCAGTTGAG 483

```


361 TTCTTCATGAAGTGAATGTTCCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT 420
 Db TTCTTCATGAAGTGAATGTTCCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT 543
 QY 421 CTTTATCTGTAAGACAGCCAAAAGTCTGCGCAATGGTGTATCTCCAGTGGATGGAG 480
 Db 544 CCTACTCTGTAAGACAGCCAAAAGTCTGCGCAATGGTGTATCTCCAGTGGATGGAG 603
 QY 481 TCAATGGAACCAATGAAGCAAGTTTATCAGATCGTGAAGCCCTTCAACCCCACTCTGTC 540
 Db 604 TCAATGGAACCAATGAAGCAAGTTTATCAGATCGTGAAGCCCTTCAACCCCACTCTGTC 663
 QY 541 TTCTTCATGAAGTGAATGTTCCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT 600
 Db 664 TTCTTCATGAAGTGAATGTTCCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT 723
 QY 601 TCGGATATCAGAAGCTTTCTGACATCCCATAGGATTTCTGGGATGATGAACAGGC 660
 Db 724 TCGGATATCAGAAGCTTTCTGACATCCCATAGGATTTCTGGGATGATGAACAGGC 783
 QY 661 ATAGCGATATCTGTGCGCGAGTGGCTCTGGGGGCGCAAGGTGTGGAACTGCACATACT 720
 Db 784 ATAGCGATATCTGTGCGCGAGTGGCTCTGGGGGCGCAAGGTGTGGAACTGCACATACT 843
 QY 721 TTGGACAGACCTGGAAGGGAGTGACACTCGGCTCGCTGAGAGCTTGAGAGACTGGCC 780
 Db 844 TTGGACAGACCTGGAAGGGAGTGACACTCGGCTCGCTGAGAGCTTGAGAGACTGGCC 903
 QY 781 GAGCTGGTGGCTGAGTGGCTCTGGGGGCGCAAGGTGTGGAACTGCACATACT 840
 Db 904 GAGCTGGTGGCTGAGTGGCTCTGGGGGCGCAAGGTGTGGAACTGCACATACT 963
 QY 841 TTGGCTGTGAGATGGCTTCAATGAGAAGCTGGGCAAGTCTGTGTGGTGGCCAAAGTGAA 900
 Db 964 TTGGCTGTGAGATGGCTTCAATGAGAAGCTGGGCAAGTCTGTGTGGTGGCCAAAGTGAA 1023
 QY 901 ATTCGGAGGAGCAGTCTTCAATGAGAAGCTGGGCAAGTCTGTGTGGTGGCCAAAGTGAA 960
 Db 1024 ATTCGGAGGAGCAGTCTTCAATGAGAAGCTGGGCAAGTCTGTGTGGTGGCCAAAGTGAA 1083
 QY 961 GCCTATCTCTGGAAGACATTTTAACTGAGTGGGCAAGTCTGTGTGGTGGCCAAAGTGAA 1020
 Db 1084 GCCTATCTCTGGAAGACATTTTAACTGAGTGGGCAAGTCTGTGTGGTGGCCAAAGTGAA 1143
 QY 1021 GAGGATGACCATCATGGAAGATTTGTTAGATAATCATGGAATAATCAAGTCTTAA 1082
 Db 1144 GAGGATGACCATCATGGAAGATTTGTTAGATAATCATGGAATAATCAAGTCTTAA 1203

RESULT 10
 BC057977
 LOCUS
 DEFINITION Mus musculus N-acetylneuraminic acid synthase (sialic acid synthase), mRNA (cdna clone MGC:67938 IMAGE:5052339), complete cds.
 ACCESSION BC057977
 VERSION BC057977.1 GI:37046731
 KEYWORDS MGC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1916)

REFERENCE
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altshul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Yax, S.I., Wang, J., Haieh, F., Diatchenko, L., Xarusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Abramson, R., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., McKernan, K.J., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.F., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, J.J., Hulyk, S.W.,

Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Rahney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Scherch, A., Schein, J.E., Jones, S.D., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

2 (bases 1 to 1916)
 Strausberg, R.
 Direct Submission
 Submitted (08-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nri.nih.gov
 Akhter, N., Avele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK plate: 123 Row: f Column: 14
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 16716466.

FEATURES

Location/Qualifiers
 1..1916
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:67938 IMAGE:5052339"
 /tissue_type="Liver, normal. 5 month old male mouse."
 /clone_lib="NCI CGAP_Li9"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1..1916
 /gene="Nans"
 /db_xref="LocusID:94181"
 /db_xref="MGI:2149820"
 5..1084
 /codon_start=1
 /product="N-acetylneuraminic acid synthase (sialic acid synthase)"
 /protein_id="AAH57977.1"
 /db_xref="GI:37046732"
 /db_xref="LocusID:94181"
 /translation="MPLELELCPCRMYGKHPFIAEIGNHQGDIDVAKRMIRTK
 ECGADCAKFKSELEKFNKALRPVTSKHSCKTGEKHELFESHDQKELQSYA
 QETGIFATGDMEMAVEFHELNPFVKVSGDTPNPFLEKTKAGKRWVSSGMQ
 SMTMKYQVAVLQPLNPFCLQCTISAYLPQDANLRVISEYQKLPFD-PIGYSGHE
 TGAISVAVALKAVLRITLTDKTKWSDHSASLEPGLAELVRSVLRVLAIGSP
 TKQLLPCENAKNEKLGKSVVAKVKKIPAGTTTLTDLMLTKVKGEKPYPEDIFNLGKK
 VLVTIEEDDTVMEESEVSHSKXKA"

gene

CDS

misc_feature		26..1057		/note="SpsE; Region: Sialic acid synthase [Cell envelope biogenesis, outer membrane]"		/db_xref="CDD:COG2089"	
ORIGIN							
Query Match		81.8%; Score 883.2; DB 10; Length 1916;					
Best Local Similarity		88.6%; Pred. No. 1.2e-207;					
Matches 957; Conservative 0; Mismatches 123; Indels 0; Gaps 0;							
QY	1	ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCGCTGGTGGGGCGGCAACACCCGCTGCTTC	60				
Db	5	ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCGCTGGTGGGGCGGCAACACCCGCTGCTTC	64				
QY	61	ATCATGTCGAGATCGGCGGAGACACACCGGCGGACCTGGAGCTAGCAAGCGCATGATC	120				
Db	65	ATCATGTCGAGATCGGCGGAGACACACCGGCGGACCTGGAGCTAGCAAGCGCATGATC	124				
QY	121	CGCATGCGCAAGGAGTGTGGGGCTGATTGTGCCAGTTCCAGAGAGTGGAGTGAATTC	180				
Db	125	CGCATGCGCAAGGAGTGTGGGGCTGATTGTGCCAGTTCCAGAGAGTGGAGTGAATTC	184				
QY	181	AAGTTTAAATCGGAAGCTTTGGAGAGCCATACCTCGAAGCATTCTCGGGGGAAGACG	240				
Db	185	AAGTTTAAATCGGAAGCTTTGGAGAGCCATATCTTCTGGAAGCATTCTCGGGGGAAGACG	244				
QY	241	TACGGGAGGACCAACGACATCTGGAGTTGACCATGACCATGACGAGGCTGAGAGG	300				
Db	245	TATGGGAGGACCAACGAGCTCTGGAATTCAGCCAGCACCAGTACAGGAGCTGAGAGC	304				
QY	301	TACGCCGAGGAGTGTGGGATCTTCTTCACTCCCTCTGCGATGATGATGATGATGATG	360				
Db	305	TATGCCGAGGAGTGTGGGATCTTCTTCACTCCCTCTGCGATGATGATGATGATGATG	364				
QY	361	TTCTGCTGACAGTGAATGTTCATTTTCAAGTTGAGTGGAGACACTAATAATTTT	420				
Db	365	TTCTGCTGACAGTGAATGTTCATTTTCAAGTTGAGTGGAGACACTAATAATTTT	424				
QY	421	CCATCTGCAAAAGACAGCAAAAGAGTGGCCCAATGGTGTGATCTCCAGTGGATGAG	480				
Db	425	CCATCTGCAAAAGACAGCAAAAGAGTGGCCCAATGGTGTGATCTCCAGTGGATGAG	484				
QY	481	TCAATGACACCATGAAAGCAAGTTTATCAGATCTGTAAGCCCTCAACCCCACTCTG	540				
Db	485	TCAATGACACCATGAAAGCAAGTTTATCAGATCTGTAAGCCCTCAACCCCACTCTG	544				
QY	541	TTCTTGAGTGTACAGCGCATCCCGCTCCAGCTGAGGACGTTCAACCTCGGGGTCATC	600				
Db	545	TTCTTGAGTGTACAGCGCATCCCGCTCCAGCTGAGGACGTTCAACCTCGGGGTCATC	604				
QY	601	TGGAATATCAGAGCTCTTTCTGACATTCCTAGGATGATGATGATGATGATGATG	660				
Db	605	TGGAATATCAGAGCTCTTTCTGACATTCCTAGGATGATGATGATGATGATGATG	664				
QY	661	ATAGCGATATCTGTGCCCGAGTGGCTCTGGGGGCGCAAGTGTGGAACCTCAATACT	720				
Db	665	ATAGCGATATCTGTGCCCGAGTGGCTCTGGGGGCGCAAGTGTGGAACCTCAATACT	724				
QY	721	TTGGACAAAGCTGGAAGGGAGTGAACCTCGGCTCGCTGAGCTGAGAACTGGCC	780				
Db	725	TTGGACAAAGCTGGAAGGGAGTGAACCTCGGCTCGCTGAGCTGAGAACTGGCC	784				
QY	781	GAGCTGGTGGCTGAGTGTCTGTGGAGCGTCCCTTGGCTCCCAACCAAGAGCTG	840				
Db	785	GAGCTGGTGGCTGAGTGTCTGTGGAGCGTCCCTTGGCTCCCAACCAAGAGCTG	844				
QY	841	CTGCCCTGTGAGATGCTCTGCAATGAGAAGCTGGGCAAGTCTGTGTGGCGCAAGTGA	900				
Db	845	CTGCCCTGTGAGATGCTCTGCAATGAGAAGCTGGGCAAGTCTGTGTGGCGCAAGTGA	904				
QY	901	ATTCGGAGGACCACTTCTACATGGAATGATGCTGACCGTGAAGTGGTGGAGCCAAA	960				
Db	905	ATTCGGAGGACCACTTCTACATGGAATGATGCTGACCGTGAAGTGGTGGAGCCAAA	964				

QY	961	GCCTATCCTCCTGAGACATCTTTAATCTAGTGGGCGAAGAGGCTCTGGTCACTGTGAA	1020				
Db	965	GGCTATCCTCCTGAGACATCTTTCAACCTAGCGGCGCAAAAGGCTCTGGTCACTATCGAA	1024				
QY	1021	GAGGATGACACCATCATGGAAGAAATGGTAGAATCATCGCAAAAAAATCAAGTCTTAA	1080				
Db	1025	GAGGATGACACCGTTCATGGAAGAAATCGTGGAAAGTCACAGCAGAAAAATCAAGGCTTAA	1084				
RESULT 11							
AB041263		1174 bp		mRNA		linear	
LOCUS		Mus musculus mRNA for N-acetylneuraminic acid 9-phosphate synthetase, complete cds.		ROD 15-JUL-2000			
ACCESSION		AB041263					
VERSION		AB041263.1					
KEYWORDS		Mus musculus (house mouse);					
SOURCE		Mus musculus					
ORGANISM		Mus musculus					
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; Nakata, D., Close, B.E., Colley, X.J., Matsuda, T. and Kitajima, K.					
AUTHORS		Molecular cloning and expression of the mouse N-acetylneuraminic acid 9-phosphate synthase which does not have deaminoneuraminic acid (KDN); 9-phosphate synthase activity					
TITLE		Biochem. Biophys. Res. Commun. 273 (2), 642-648 (2000)					
JOURNAL		20334323					
MEDLINE		10873658					
PUBMED		2 (bases 1 to 1174)					
REFERENCE		Kitajima, K. and Nakata, D.					
AUTHORS		Submitted (04-APR-2000) Ken Kitajima, Graduate school of bioagricultural sciences, Nagoya University, Department of applied Molecular biosciences; Chikusa-ku Huroo-chou, Nagoya, Aichi 464-8601, Japan (E-mail: kitajima@agr.nagoya-u.ac.jp).					
JOURNAL		Tel: +81-52-789-4130, Fax: +81-52-789-4126)					
TITLE		Location/Qualifiers					
FEATURES		1..1174					
source		/organism="Mus musculus"					
		/mol_type="mRNA"					
		/db_xref="taxon:10090"					
		/cell_line="COMGid"					
		71..11150					
		/codon_start=1					
		/product="N-acetylneuraminic acid 9-phosphate synthetase"					
		/protein_id="BAA98131.1"					
		/db_xref="GI:8978234"					
		/translation="MPLELECPGRVWGKHPFIIAIFGIMQHSDIDVAXEMIRTAKEGADAFQKSELEFENRKALRDPYTSKHSWGTNYGEHRHLFSFQKVKELQSYA QEIIGFTFAKDEMAVEFHELVNPFYKVGSGDTNPNFYLEKTAQGRPMVIGSQW SMDTKOVYQIVKPLNFCFLQCTSAVPLQPEDANLRVISEYOKLPDIPDIGYSGHE TGIASVAAVAIGAKVLERHITLDTKWSGSDHSALEGEALAEALRSVRLVERALGSP TKOLLPEMACNEKLGSVYAKVXKIPAGTILTLTLTLVVKVGEPEKIPPEDFNLAGKPVLTIEEDDTVMESVESHSKKIKA"					
CDS							
Query Match		81.6%; Score 881.6; DB 10; Length 1174;					
Best Local Similarity		88.5%; Pred. No. 2.9e-207;					
Matches 956; Conservative 0; Mismatches 124; Indels 0; Gaps 0;							
QY	1	ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCGCTGGTGGGGCGGCAACACCCGTCGTC	60				
Db	71	ATGCCGCTGGAACTGGAGCTGTGTCTCCGGGGCGCTGGTGGGGCGGCAACACCCGTCGTC	130				
QY	61	ATCATTCGCCGAGATCGGCGGAGACCCACGAGGACCTGGACGTAGCCAGCGCATGATC	120				
Db	131	ATCATTCGCCGAGATCGGCGGAGACCCACCAAGGAGACATAGATGTGGCGCAAGCGCATGATC	190				
QY	121	CGCATGCCCAAGAGTGTGGGGCTGATTGTGCCAAGTTCCAGAGAGCTGAGCTAGATTC	180				
Db	191	CGCATGCCCAAGAGTGTGGGGCTGATTGTGCCAAGTTCCAGAGAGCTGAGCTAGATTC	180				

QY 961 GCCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGAGTCTCTGGTCACTGTTGA 1020
 DB 965 GCCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGAGTCTCTGGTCACTGTTGA 1024
 QY 1021 GAGGATGACACCATCATGGAAGAAATGTTGATGATAATCATGGCAAAAATCAAGTCTTAA 1080
 DB 1025 GAAGATGACACCGTTCATGGAAGAAATCCGTGGAAGTACACGCAAGAAATCAAGGCTTAA 1084

RESULT 11
 AB041263 1174 bp mRNA linear ROD 15-JUL-2000
 LOCUS Mus musculus mRNA for N-acetylneuraminic acid 9-phosphate
 DEFINITION synthetase, complete cds.
 ACCESSION AB041263
 VERSION AB041263.1 GI:8978233
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus

REFERENCE 1 (sites) Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS Nakata,D., Close,B.E., Colley,K.J., Matsuda,T. and Kitajima,K.
 TITLE Molecular cloning and expression of the mouse N-acetylneuraminic acid 9-phosphate synthase which does not have deaminoneuraminic acid (KDN): 9-phosphate synthase activity
 JOURNAL Biochem. Biophys. Res. Commun. 273 (2), 642-648 (2000)
 MEDLINE 20334323
 PUBMED 10873658

REFERENCE 2 (bases 1 to 1174) Kitajima,K. and Nakata,D.
 TITLE Direct Submission
 JOURNAL Submitted (04-APR-2000) Ken Kitajima, Graduate school of bioagricultural sciences,Nagoya University, Department of applied Molecular biosciences; Chikusa-ku Hurou-chou, Nagoya, Aichi 464-8601, Japan (E-mail:kitajima@agr.nagoya-u.ac.jp, Tel:+81-52-789-4130, Fax:+81-52-789-4128)
 FEATURES Location/Qualifiers
 source 1..1174
 organism="Mus musculus"
 mol_type="mRNA"
 db_xref="taxon:10090"
 cell_line="COWAld"
 71..1150
 /codon_start=1
 /product="N-acetylneuraminic acid 9-phosphate synthetase"
 /protein_id="BA098131.1"
 /db_xref="GI:8978234"

CDs
 1. 1174
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /cell_line="COWAld"
 71..1150
 /codon_start=1
 /product="N-acetylneuraminic acid 9-phosphate synthetase"
 /protein_id="BA098131.1"
 /db_xref="GI:8978234"
 /translation="MPLELECPGRWVGKHPFIIAEIQNHOGDIDVAKEMRTAK
 ECGADCAKFOKSELEFKNRKALEPYTSKHSWGTGHEKRLHLSFSDQYKELQSYA
 QETGIFTAGMDMAVEFLHNLNVPFFKVGSDTANNFVLEKTKAGRMVSSGMQ
 SMDTKOVYIVKPLNPFCLQCTSAFPLQPEDANLRVISEYKLPDPIDPIGYSGHE
 TGAISVAVALGAKVLERHITLDTKWDSDHSAJEPGELAEIVSRVIRALGSP
 TKLLPCEWACNEKUGKSVVAKVIPAGTTLTLDTLTVKVGEPKXPEDIFNLGKK
 VLVTIEDPTVMEESVESHSKKIKA"

ORIGIN
 Query Match 81.6%; Score 881.6; DB 10; Length 1174;
 Best Local Similarity 88.5%; Pred. No. 2.9e-207;
 Matches 956; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
 QY 1 ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCGCTGGTGGGGCGGCAACACCCGCTGCTTC 60
 DB 71 ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCGCTGGTGGGGCGGCAACACCCGCTGCTTC 130
 QY 61 ATCATGTCGAGATCGGCGGAGACACACCGGCGGACCTGGAGCTAGCAAGCGCATGATC 120
 DB 131 ATCATGTCGAGATCGGCGGAGACACACCGGCGGACCTGGAGCTAGCAAGCGCATGATC 190
 QY 121 CGCATGCGCAAGGAGTGTGGGGCTGATTGTGCCAGTTCCAGAGAGTGGAGTGAATTC 180
 DB 191 CGCATGCGCAAGGAGTGTGGGGCTGATTGTGCCAGTTCCAGAGAGTGGAGTGAATTC 250

QY	181	AGTTTAAATCGAAAGCCCTTGGAGAGCCCATACACTCGAAGCATTTCTCTGGGGGAAGACG	240	
DB	251	AAGTTTAAACCGGAAGCCCTCGAGAGACCATATACTTCGAAGCATTCATCGGGAGAGCG	310	
QY	241	TACGGGGAGCACAACGACATCTGAGATTCAGCCTACAGCAGTACAGGAGCTGCAGAGG	300	
DB	311	TATGGGGAGCACAACGCGCATCTGAAATTCAGCCACGACCCAGTACAAAGAGAGCTGCAGAGC	370	
QY	301	TACGCCAGAGGAGTTGGGATCTTCTTCACTGCTCTGGCATGAGTGCAGATGCACTTCGAA	360	
DB	371	TATGCGCAGGAGATCGGCATCTTCTTCACTGCTCTGSCATGSAAGATGGCAGTTGAG	430	
QY	361	TTCTCTGCTGAACTGAATGTTCCATTTTCAAAAGTTGGATCTGAGAGACACTAAATAATTTT	420	
DB	431	TTTCTGCAAGCACTGAATGTTTCCCTTTTCAAGTTGGATCTGAGGACACTAAACAATTTT	490	
QY	421	CCCTATCTGGAAAAGACAGCAAAAAGSTGCCCACTGGTATCTCCAGTGGGATGCGAG	480	
DB	491	CCCTACCTGGAAAAGACAGCCAAAGAGTGTCTTATGTTGATCTCCAGCGGGATGCGAG	550	
QY	481	TCATTTGACACCATGAAGCAAGTTTATCAGATCTGTGAAGCCCTTCAACCCCAACTTCTGCG	540	
DB	551	TCAATTTGACACCATGAAGCAAGTCTATCAGATCTGTGAAGCCCTGATATCCCACTTCTGCG	610	
QY	541	TTCTTGTGAGTATCAGCGGCATACCCGCTTCAGCTTGAGGACGTCAAGCTCGGGGTCTATC	600	
DB	611	TTCTCTCCAATGCACAGCGCGTACCCACTACAGCCGAGGATGCAACCTCGCGGTCTATC	670	
QY	601	TCGGAATATCAGAAAGCTCTTTCTTGACATTCCTATAGGTTATCTGAGCATGAACAGGC	660	
DB	671	TCGGAATACAGAAAGCTCTTTCCGACATTCCTATCGGTTATTCGGGCACAGACGGCG	730	
QY	661	ATACGGATATCTGTGGCGCATGTGGCTCTGGGGCCCAAGTGTGTGGAACGTCACTAACT	720	
DB	731	ATCCGCCATATCTGTGGCGCGCTGGCTCTGGGGCCCAAGTGTGTGGAACGTCACTAACT	790	
QY	721	TTGACACAAGACCTGGAAGGGAGTGAACCATCGGCTCTGGGGCCCAAGTGTGTGGAACGTCACTAACT	780	
DB	791	TTGACACAAGACCTGGAAGGGAGTGAACCATCGGCTCTGGGGCCCAAGTGTGTGGAACGTCACTAACT	850	
QY	781	GAGCTGTGGGTCAAGTGGCTTTGTGGAGCGTCCCTTGGGCTCCCAACCAAGCAGCTG	840	
DB	851	GAGCTGTGGGTCAAGTGGCTTTGTGGAGCGTCCCTTGGGCTCCCAACCAAGCAGCTG	910	
QY	841	CTGCGCTGTGAGATGGCTGCAATGAGAGTGGGCAAGTCTGTGTGGCCCAAGTGAAG	900	
DB	911	CTGCGCTGTGAGATGGCTGCAATGAGAGTGGGCAAGTCTGTGTGGCCCAAGTGAAG	970	
QY	901	ATTCGGGAAGCACAATCTTAACAATGACATCTCACTGAGAGTGGGTGAGCCCAAA	960	
DB	971	ATCCAGCAGCACCACCTTGACCTTGACACGCTCACTGTGAGTGGGGAGCCCAAA	1030	
QY	961	GCCTATCTCTCTGAAGACATCTTTAATCTAGTGGGCAAGAGTCTCTGTGCTATGTTGAA	1020	
DB	1031	GGCTATCTCTCTGAAGACATCTTTAATCTAGTGGGCAAGAGTCTCTGTGCTATGTTGAA	1090	
QY	1021	GAGATGACACCATCATGGAAGATTTGATAGATATCTGCGCAAAAATCAAGTCTTAA	1080	
DB	1091	GAAGATGACACCATCATGGAAGATTTGATAGATATCTGCGCAAAAATCAAGTCTTAA	1150	
RESULT 12	AK091184	3026 bp	linear	PR-15-JUL-2002
LOCUS	AK091184			
DEFINITION	Homo sapiens cDNA FLJ33865 fis, clone CTONG2006515, highly similar to N-acetylneuraminic acid phosphate synthase.			
ACCESSION	AK091184			
VERSION	AK091184.1	GI:21749493		
KEYWORDS	cligo capping; fis (full insert sequence).			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			


```

Db 71 ATCCGCTGGAGCTCGAGCTGTGTCCCGGCGCTGGTGGCGGCAACACCCGCTCTC 130
QY 61 ATCATTCGGGATCGGCGCAGAACACACAGGCGGACCTGGAGTASCCAGCGATGATC 120
Db 131 ATCATTCGGGATCGGCGCAGAACACACAGGCGGACCTGGAGTASCCAGCGATGATC 190
QY 121 CGCATGGCCAGGAGTGTGGGCTGATTTGTGCAAGTTCACAGAGAGTGAAGTATTC 180
Db 191 CGCATGGCCAGGAGTGTGGGCTGATTTGTGCAAGTTCACAGAGAGTGAAGTATTC 250
QY 181 AAGTTTATCGGAAGCCCTTGGAGAGGCATACACCTCGAAGCATTCCTGGGGGAGAGC 240
Db 251 AAGTTTATCGGAAGCCCTTGGAGAGGCATACACCTCGAAGCATTCCTGGGGGAGAGC 310
QY 241 TACGGGGAGCACACAGCATCTGGAGTTCAGCCATGACAGGAGTGAAGAGTGAAGC 300
Db 311 TACGGGGAGCACACAGCATCTGGAGTTCAGCCATGACAGGAGTGAAGAGTGAAGC 370
QY 301 TACGGGGAGGAGTGGGATCTCTTCACTGCTCTGGCATGATGAGATGSCAGTTCMA 360
Db 371 TACGGGGAGGAGTGGGATCTCTTCACTGCTCTGGCATGATGAGATGSCAGTTCMA 430
QY 361 TTCTGCTGATGAGTGTTCATTTTCAAGTTGGATCTGGAGACACTAATATTTT 420
Db 431 TTCTGCTGATGAGTGTTCATTTTCAAGTTGGATCTGGAGACACTAATATTTT 490
QY 421 CCTTATCTGGAAGAGCAGCAAAAGGTCGCCAATGGTGTCTCCAGTGGATGAG 480
Db 491 CCTTATCTGGAAGAGCAGCAAAAGGTCGCCAATGGTGTCTCCAGTGGATGAG 550
QY 481 TCAATGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCTGC 540
Db 551 TCAATGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCTGC 610
QY 541 TTCTGCTGATGAGTGTTCATTTTCAAGTTGGATCTGGAGACACTAATATTTT 600
Db 611 TTCTGCTGATGAGTGTTCATTTTCAAGTTGGATCTGGAGACACTAATATTTT 670
QY 601 TCGGAATATCAGAGCTCTTCTGACATTCCTGAGTGTTCGAGTGTTCGAGTGTTC 660
Db 671 TCGGAATATCAGAGCTCTTCTGACATTCCTGAGTGTTCGAGTGTTCGAGTGTTC 730
QY 661 ATACGATATCTGGGCGCAGTGGC-TCTGGGGCCCAAGGTGTGGAACTGCATAAC 719
Db 731 ATACGATATCTGGGCGCAGTGGC-TCTGGGGCCCAAGGTGTGGAACTGCATAAC 789
QY 720 TTGGCAAGACCTGGAGGGAG 743
Db 790 TTGGCAAGACCTGGAGGGAG 813

```

```

RESULT 15
AX873806/c
LOCUS AX873806 514 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 8711 from Patent EP1074617.
ACCESSION AX873806
VERSION AX873806.1 GI:40028595
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 8711 07-FEB-2001;
RESEARCH Association for Biotechnology (JP)
FEATURES
source .. 514
/organism="Homo sapiens"
/mol_type="unassigned DNA"

```

```

ORIGIN
/db_xref="taxon:9606"
Query Match 44.8%; Score 483.8; DB 6; Length 514;
Best Local Similarity 98.2%; Pred. No. 8.6e-109;
Matches 485; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 587 ACCTGGGGTCACTCGGATATCAGAGCTCTTTCTTCTGACATTCCTCATAGGGTATCTGG 646
Db 514 ACNTGGGGTNNNTNGGAATATCAGAACTCTTTCTTCTGACATTCCTCATAGGGTATCTGG 455
QY 647 GGCATGAACACAGCCTATAGCATATCTGTGGCCGCGAGTGGCTCTGGGGGCAAGGTGTGG 706
Db 454 GGCATGAACACAGCCTATAGCATATCTGTGGCCGCGAGTGGCTCTGGGGGCAAGGTGTGG 395
QY 707 AACGTCAATCACTTTGGACAGACCTGGAAAGGAGTGAACCACTCGGCTCGCTGGAGC 766
Db 394 AACGTCAATCACTTTGGACAGACCTGGAAAGGAGTGAACCACTCGGCTCGCTGGAGC 335
QY 767 CTGGAGAACTGGCCGAGCTGGTGGCTCAGTCCGCTCTTGTGGAGCGTGGCTGGCTGCC 826
Db 334 CTGGAGAACTGGCCGAGCTGGTGGCTCAGTCCGCTCTTGTGGAGCGTGGCTGGCTGCC 275
QY 827 CAACCAAGCAGCTGCTGCCCTGTGAGATGGCTGCATGAGAAAGCTGGGCAAGTCTGTGG 886
Db 274 CAACCAAGCAGCTGCTGCCCTGTGAGATGGCTGCATGAGAAAGCTGGGCAAGTCTGTGG 215
QY 887 TGGCCAAAGTGAATAATTCGGAGAGGCACTTCTTAACATGGAATGATCTACCGTGAAG 946
Db 214 TGGCCAAAGTGAATAATTCGGAGAGGCACTTCTTAACATGGAATGATCTACCGTGAAG 155
QY 947 TGGGTGAGCCCAAGCCCTATCTCTGAGAGACATCTTTAATCTAGTGGGCAAGAGGTCC 1006
Db 154 TGGGTGAGCCCAAGCCCTATCTCTGAGAGACATCTTTAATCTAGTGGGCAAGAGGTCC 95
QY 1007 TGGTCACTGTTGAAGAGGATGACACCATCATGGAAGAAATGGTAGATAATCATGGCAAA 1066
Db 94 TGGTCACTGTTGAAGAGGATGACACCATCATGGAAGAAATGGTAGATAATCATGGCAAA 35
QY 1067 AAATCAAGTCTTAA 1080
Db 34 AAATCAAGTCTTAA 21

```

Search completed: May 11, 2004, 17:15:36
Job time : 4392.16 Secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2004, 04:14:30 ; Search time 465.643 Seconds
(without alignments)
9853.151 Million cell updates/sec

Title: US-09-930-440B-5

Perfect score: 1080

Sequence: 1 atgcgcgtggagctggagct.....gcaaaaaaatcaagtcttaa 1080

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124059041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N Geneseq_29Jan04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002s.*
7: Geneseqn2003as.*
8: Geneseqn2003bs.*
9: Geneseqn2003cs.*
10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1080	100.0	1080	3	AAA37763 Human gly
2	1080	100.0	1080	4	AAB28458 Nucleotid
3	1080	100.0	1080	7	AAL53993 DNA encod
4	1080	100.0	1080	3	AAA50569 Human sia
5	1078	99.8	1268	6	ABL90630 Human pol
6	1076.8	99.7	1230	4	AAI59142 Human pol
7	1076.8	99.7	1230	8	ADB49125 Novel hum
8	1075.2	99.6	1170	4	AAH14557 Human CDN
9	998.2	83.2	1243	4	AAI60928 Human pol
10	732	67.8	1238	4	AAI60928 Human pol
11	710.2	65.8	813	4	AAH06847 Human CDN
12	483.8	44.8	514	4	AAH11876 Human CDN
13	426.4	39.5	1417	5	AAI59142 Human pol
14	418	38.7	454	6	ABN95016 Gene #151
15	386.8	35.8	553	6	ABK94731 Neurodege
16	371.8	34.4	486	8	ACH18008 Human adu
17	336.8	31.2	442	8	ACH1408 Human fo
18	327.8	30.4	417	8	ACH38131 Human end
19	315.8	29.2	480	8	ACH32141 Human end
20	280.6	26.0	404	5	AAH42798 Nucleotid
21	260.6	24.1	1028	5	AAH81438 DNA encod
22	256.2	23.7	410	4	AAH35557 Human col
23	198.2	18.4	1017	4	ABL29095 Drosophil

24	162.4	15.0	611	5	AA591836 DNA encod
25	146	13.5	492	4	AAK88570 Human dig
26	119.8	11.1	3123	4	ABL29094 Drosophil
27	119.8	11.1	3375	4	ABL29196 Drosophil
28	112.4	10.4	26281	3	AAZ60929 Nucleotid
29	101.6	9.4	17276	7	ACA64723 Streptoco
30	100	9.3	1023	6	ABN67441 Streptoco
31	100	9.3	110000	6	ABN71527 11
32	90	8.3	3777	4	ABL29200 Drosophil
33	70.4	6.5	110000	2	AAV21209 04
34	63	5.8	1059	3	AA50570 Escherich
35	63	5.8	1059	4	AAH28459 Nucleotid
36	63	5.8	1059	7	AAU53994 DNA encod
37	62.8	5.8	999	4	AA591405 N. mening
38	62.8	5.8	1000	6	ABK37785 DNA seque
39	62.8	5.8	1050	6	ABA99765 N. mening
40	62.8	5.8	65632	3	AAH81502 N. mening
41	62.8	5.8	110000	3	AAH81490 00
42	62.8	5.8	349980	3	AAF21544 Neisseria
43	62	5.7	133	3	AAC98691 Human col
44	60	5.6	60	6	ABN36327 Human spl
45	59.6	5.5	984	6	ABV72354 Nucleotid

ALIGNMENTS

RESULT 1

AAA37763

ID AAA37763 standard; DNA; 1080 BP.

XX AAA37763;

AC AAA37763;

DT 04-DEC-2000 (first entry)

XX Human glycosylation enzyme clone HASAA37 coding sequence.

XX Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
immunotherapy; cosmetic surgery; metabolism; immune system disorder;
haematopoietic cell deficiency; blood coagulation disorder; asthma;
afibrinogenemia; blood platelet disorder; thrombocytopenia; necropia;
autoimmune disorder; Addison's disease; multiple sclerosis; neoplasia;
allergic encephalomyelitis; allergic reaction; organ rejection;
graft-versus-host disease; inflammation; hyperproliferative disorder;
sarcoidosis; infection; gene therapy; CMP sialic acid synthetase; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
CDS 1..1080
/*tag= a
/product= "Sialic acid synthetase"

WO2000052136-A2.

PD 08-SEP-2000.

PF 01-MAR-2000; 2000WO-US005325.

PR 02-MAR-1999; 99US-0122409P.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Coleman TA;

XX WPI: 2000-572179/53.

DR P-PSDB; AAY90352.

XX New human glycosylation enzymes cytidine 5'-monophosphate sialic acid
synthetase, sialic acid synthetase and aldolase and nucleic acids
encoding the proteins for treating e.g., immune system disorders,
microbial diseases.

PS Disclosure; Page 103-105; 144pp; English.

CC The present sequence is that of human cDNA encoding sialic acid
 CC synthetase (see AY96101), an enzyme that condenses MannAC-6-P or Man-6-P
 CC with PP to form Neu5Ac and KDN phosphates, respectively. This novel
 CC gene, termed SAS, was identified on the basis of homology with the
 CC Escherichia coli neuB gene using a bioinformatics based approach.
 CC Northern blots indicated ubiquitous transcription of the human synthetase
 CC gene in a selection of tissues. The invention provides methods and
 CC recombinantly engineered cells for producing glycoproteins having
 CC sialylated oligosaccharides. The methods involve altering the expression
 CC of enzymes involved in carbohydrate processing. A claimed cell producing
 CC sialylated glycoprotein at above endogenous levels expresses at least 1
 CC (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme
 CC catalyzing the conversion of UDP-GlcNAc to YnnNAc, sialic acid
 CC synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid
 CC transporter at above endogenous levels. Endogenous N-
 CC acetylglucosaminidase activity may be suppressed. A claimed method for
 CC manipulating glycoprotein in an insect cell comprises enhancing the
 CC expression of 1 of the above enzymes, and a claimed method for producing
 CC sialylated glycoproteins involves expressing a heterologous protein
 CC (especially plasminogen, transferrin, Na,K-ATPase or thyrotropin) in
 CC the insect cell. Yeast, insect, fungal, plant and bacterial host cells
 CC can be engineered to produce new forms of sialylated glycoproteins,
 CC higher concentrations of sialylated glycoproteins and/or elevated
 CC concentrations of donor substrates (e.g. nucleotide sugars) required for
 CC sialylation
 CC
 CC

SQ Sequence 1146 BP; 290 A; 280 C; 332 G; 244 T; 0 U; 0 Other;

Query Match 100.0%; Score 1080; DB 3; Length 1146;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCTGGAGCTGGAGCTGTGTCCTGGGGGCTGGTGGGGGCGGCAACACCGCTGCTTC 60
 Db ATGCGCTGGAGCTGGAGCTGTGTCCTGGGGGCTGGTGGGGGCGGCAACACCGCTGCTTC 105
 Qy 61 ATCATTCGCGAGTCCGAGACACACACGAGGCGACCTGGAGCTAGCCAGGCGATGATC 120
 Db ATCATTCGCGAGTCCGAGACACACACGAGGCGACCTGGAGCTAGCCAGGCGATGATC 165
 Qy 121 GCGATGCGCAAGGAGTGTGGGCTGATTGTCCAAAGTTCGAGAGAGTGAAGTTC 180
 Db GCGATGCGCAAGGAGTGTGGGCTGATTGTCCAAAGTTCGAGAGAGTGAAGTTC 225
 Qy 181 AAGTTTAAATCGGAAGCTTTGGAGAGGCGATACACCTCGAAGCAATTCCTGGGGGAAGCG 240
 Db AAGTTTAAATCGGAAGCTTTGGAGAGGCGATACACCTCGAAGCAATTCCTGGGGGAAGCG 285
 Qy 241 TAGGGGAGCACACACGACATCTCGAGTTCAGCCATGACAGTACAGGAGCTGCAGAGG 300
 Db TAGGGGAGCACACACGACATCTCGAGTTCAGCCATGACAGTACAGGAGCTGCAGAGG 345
 Qy 301 TAGCCGAGGAGGTGGGATCTTCTTCACTGCTCTGGCATGGATGAGATGCGAGTTGAA 360
 Db TAGCCGAGGAGGTGGGATCTTCTTCACTGCTCTGGCATGGATGAGATGCGAGTTGAA 405
 Qy 361 TTCTGTGATGAATGATGTTCCATTTTCAAGTTGATCTGGAGACACTAATAATTTT 420
 Db TTCTGTGATGAATGATGTTCCATTTTCAAGTTGATCTGGAGACACTAATAATTTT 465
 Qy 421 CCTTATCTGAAAGAGACGCAAAAAGTCCGCAATGCTGATCTCCAGTGGGATCGAG 480
 Db CCTTATCTGAAAGAGACGCAAAAAGTCCGCAATGCTGATCTCCAGTGGGATCGAG 525
 Qy 481 TCAATGACACCATGAAAGCAAGTTTATCAGATCTGTGAAGCCCTCAACCCCACTCTGCG 540
 Db TCAATGACACCATGAAAGCAAGTTTATCAGATCTGTGAAGCCCTCAACCCCACTCTGCG 585
 Qy 541 TTCTTGAGTGTACAGCGCATCCCGCTCCAGCTGAGGACGTCACACCTGCGGGTCAATC 600
 Db TTCTTGAGTGTACAGCGCATCCCGCTCCAGCTGAGGACGTCACACCTGCGGGTCAATC 645

Qy 601 TCGGAATATCAGAAGCTCTTTCTGACATTCCTCATAGGGTATTTCTGGGCATGAAACAGGC 660
 Db TCGGAATATCAGAAGCTCTTTCTGACATTCCTCATAGGGTATTTCTGGGCATGAAACAGGC 705
 Qy 661 ATAGCGATATCTGTGGCCCGCAGTGGCTCTGGGGGCGCAAGGTGTGTGAACTCACTAAT 720
 Db ATAGCGATATCTGTGGCCCGCAGTGGCTCTGGGGGCGCAAGGTGTGTGAACTCACTAAT 765
 Qy 721 TTGGCAACAGCTCTGAAAGGCGAGTGACCACTCGGCTCGCTGGAGCTTGAGAACTGGCC 780
 Db TTGGCAACAGCTCTGAAAGGCGAGTGACCACTCGGCTCGCTGGAGCTTGAGAACTGGCC 825
 Qy 781 GAGCTGTGGCTGAGTGGCTCTTTGTGGAGCGTGGCTCGGCTCCCAACCAAGCAGCTG 840
 Db GAGCTGTGGCTGAGTGGCTCTTTGTGGAGCGTGGCTCGGCTCCCAACCAAGCAGCTG 885
 Qy 841 CTGGCTGTGGAGTGGCTCGCTGCAATGAGAGCTGGGCAAGTCTGTGTGGCCAAAGTAA 900
 Db CTGGCTGTGGAGTGGCTCGCTGCAATGAGAGCTGGGCAAGTCTGTGTGGCCAAAGTAA 945
 Qy 901 ATTCGGAGGCGACCAATCTTAACAATGGACATGCTCACCGTCAAGGTGGTGGCCCAAA 960
 Db ATTCGGAGGCGACCAATCTTAACAATGGACATGCTCACCGTCAAGGTGGTGGCCCAAA 1005
 Qy 961 GCCTATCTCTCTGAAAGACATCTTTAATCTAGTGGGCGAAGGTCCTGCTCACTGTGAA 1020
 Db GCCTATCTCTCTGAAAGACATCTTTAATCTAGTGGGCGAAGGTCCTGCTCACTGTGAA 1065
 Qy 1021 GAGGATGACCATCATCTGGAAGATTTGGTAGATAATCATGGCAAAAATCAAGTCTTAA 1080
 Db GAGGATGACCATCTGGAAGATTTGGTAGATAATCATGGCAAAAATCAAGTCTTAA 1125

RESULT 5

ABL90630
 ID ABL90630 standard; cDNA; 1268 BP.
 XX ABL90630;
 AC ABL90630;
 DT 24-MAY-2002 (first entry)
 XX Human polynucleotide SEQ ID NO 1192.
 DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; gene; ss.
 XX Homo sapiens.
 OS WC2001:90304-A2.
 FN 29-NOV-2001.
 PD 18-MAY-2001; 2001WO-US016450.
 PF 19-MAY-2000; 2000US-0205515P.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA Birse CE, Rosen CA;
 PI WPI; 2002-122018/16.
 DR P-PSDB; ABB90221.
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders.

PS Claim 4; SEQ ID NO 1192; 2081pp + Sequence Listing; English.

Query Match		99.7%; Score 1075.8; DB 8; Length 1230;
Best Local Similarity		99.8%; Pred. No. 5.2e-314;
Matches 1075; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
QY	1	ATGCCCTGGAGCTGGAGCTGTCTCCGGCGCTGGGTGGCGGGCAACACCGTCTTC 60
DB	119	ATGCCCTGGAGCTGGAGCTGTCTCCGGCGCTGGGTGGCGGGCAACACCGTCTTC 178
QY	61	ATCATTTCCGAGATCGCCAGAACACACCGAGGCGCTGACCTAGCCAGCCATGATC 120
DB	179	ATCATTTCCGAGATCGCCAGAACACACCGAGGCGCTGACCTAGCCAGCCATGATC 238
QY	121	CGCATGCCCAAGAGTGTGGGGCTGATTTGTGCCAAGTTCAGAGAGTGTAGTATTC 180
DB	239	CGCATGCCCAAGAGTGTGGGGCTGATTTGTGCCAAGTTCAGAGAGTGTAGTATTC 298
QY	181	AAATTATTCGGAAGCCTTGGAGAGCCATACACCTCGAAGCATTTCTGGGGAGAGC 240
DB	299	AAATTATTCGGAAGCCTTGGAGAGCCATACACCTCGAAGCATTTCTGGGGAGAGC 358
QY	241	TACGGGGAGCACAAAGCATCTGGAGTTTCAGCCATGACAGTACAGGAGCTGCAGAG 300
DB	359	TACGGGGAGCACAAAGCATCTGGAGTTTCAGCCATGACAGTACAGGAGCTGCAGAG 418
QY	301	TACGGGGAGGAGTTGGGATCTTTCTCACTGCTCTTSCATGATGAGATGCAATTGAA 360
DB	419	TACGGGGAGGAGTTGGGATCTTTCTCACTGCTCTTSCATGATGAGATGCAATTGAA 478
QY	361	TTCTTGCATGACTGATGTTTCAATTTTCAAGTTGGATCTGGAGACTATATATTT 420
DB	479	TTCTTGCATGACTGATGTTTCAATTTTCAAGTTGGATCTGGAGACTATATATTT 538
QY	421	CTTTATCTGGAAAGACACCAAAAGGTTCGCCCAATGCTGATCTCCAGTGGGATGAG 480
DB	539	CTTTATCTGGAAAGACACCAAAAGGTTCGCCCAATGCTGATCTCCAGTGGGATGAG 598
QY	481	TCATGGACACCAAGAGCAAGTTTATCATGATGTTGAAGCCCTCAACCCCACTCTCC 540
DB	599	TCATGGACACCAAGAGCAAGTTTATCATGATGTTGAAGCCCTCAACCCCACTCTCC 658
QY	541	TTCTTGCATGACTGATGTTTCAATTTTCAAGTTGGATCTGGAGACTATATATTT 600
DB	659	TTCTTGCATGACTGATGTTTCAATTTTCAAGTTGGATCTGGAGACTATATATTT 718
QY	601	TGCAATATCAGAGCTCTTTCTGATTCATTCATAGGTTATCTGGGATGAAACAGGC 660
DB	719	TGCAATATCAGAGCTCTTTCTGATTCATTCATAGGTTATCTGGGATGAAACAGGC 778
QY	661	ATAGCGATATCTGTGGCGCAGTGGCTCTGGGGGCCAAGGTGTGGAAGTCACTAACT 720
DB	779	ATAGCGATATCTGTGGCGCAGTGGCTCTGGGGGCCAAGGTGTGGAAGTCACTAACT 838
QY	721	TTGGACAGAGCTGGAAGGGAGTGCACCTCTGGCTCTGAGGCTTGGAGTGGCTGGC 780
DB	839	TTGGACAGAGCTGGAAGGGAGTGCACCTCTGGCTCTGAGGCTTGGAGTGGCTGGC 898
QY	781	GAGCTGGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAG 840
DB	899	GAGCTGGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAG 958
QY	841	CTGCGCTGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGGTGGCCAAAGTAA 900
DB	959	CTGCGCTGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGGTGGCCAAAGTAA 1018
QY	901	ATTCGGAGGACCATCTTCAATGAGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAG 960
DB	1019	ATTCGGAGGACCATCTTCAATGAGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAG 1078
QY	961	GCCTATCTCTTGAAGACATCTTTAATCTAGTGGGCAAGAGGTCTGTGCTACTGTGAA 1020
DB	1079	GCCTATCTCTTGAAGACATCTTTAATCTAGTGGGCAAGAGGTCTGTGCTACTGTGAA 1138
QY	1021	GAGGATGACCATCATGGAAGATTTGGTAGAATATCATGCGAAGAAATCAAGTCTTAA 1080

Db 1139 GAGGATGACCATCATGGAAGATTTGGTAGAATATGCGCAAAAAATCAAGTCTTAA 1198

RESULT 8
AAH14557
ID AAH14557 standard; cDNA; 1170 BP.
XX
AC AAH14557;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12129.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PK EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 12129; 2537bp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95993 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

Sequence 1170 BP; 294 A; 285 C; 347 G; 244 T; 0 U; 0 Other;

Query Match 99.6%; Score 1075.2; DB 4; Length 1170;
Best Local Similarity 99.7%; Pred. No. 1.6e-313;
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGCGCTGGAGCTGGAGCTGTGTCCCGGGCGCTGGTGGTGGCGGCAACACCCGCTGCTC 60
Db 71 ATGCGCTGGAGCTGGAGCTGTGTCCCGGGCGCTGGTGGTGGCGGCAACACCCGCTGCTC 130
QY 61 ATCATTTCCGAGATCGCCGAGAACCCACCGGCGACCTGGACGTAGCCAAAGCGATGATC 120
Db 131 ATCATTTCCGAGATCGCCGAGAACCCACCGGCGACCTGGACGTAGCCAAAGCGATGATC 190
QY 121 CGCATGCCAAGGAGTGGGGCTGATTGTCGAAGTTCCAGAAAGAGTGAGCTAGAAATTC 180
Db 191 CGCATGCCAAGGAGTGGGGCTGATTGTCGAAGTTCCAGAAAGAGTGAGCTAGAAATTC 250
QY 181 AAGTTTAAATCGAAAGCTTGGAGAGCCATACACCTCGAAGCAATCTCTGGGGGAAGACG 240
Db 251 AAGTTTAAATCGAAAGCTTGGAGAGCCATACACCTCGAAGCAATCTCTGGGGGAAGACG 310
QY 241 TACGGGAGCACAACGACATCTGGAGTTCAGCCATGACCAAGTACAGGAGCTGCAGAGG 300
Db 311 TACGGGAGCACAACGACATCTGGAGTTCAGCCATGACCAAGTACAGGAGCTGCAGAGG 370
QY 301 TACGCCGAGAGGTTGGGATCTTCTTCACTGCCTCTGSCATGGATGAGATGSCAATTCGA 360
Db 371 TACGCCGAGAGGTTGGGATCTTCTTCACTGCCTCTGSCATGGATGAGATGSCAATTCGA 430
QY 361 TTCTCGATGAACTAAGTTGCAATTTTCAAGTTGAGTCTGAGAGACACTAAATATTT 420
Db 431 TTCTCGATGAACTAAGTTGCAATTTTCAAGTTGAGTCTGAGAGACACTAAATATTT 490
QY 421 CCTTATCTGGAAGAGCAGCCAAAAGGTGCGCCAAATGGTGATCTCCAGTGGGATCGAG 480
Db 491 CCTTATCTGGAAGAGCAGCCAAAAGGTGCGCCAAATGGTGATCTCCAGTGGGATCGAG 550
QY 481 TCAATGACACCATGAAGCAAGTTTATCAGATCTGTGAGGCCCTCAACCCCAACTTCTGC 540
Db 551 TCAATGACACCATGAAGCAAGTTTATCAGATCTGTGAGGCCCTCAACCCCAACTTCTGC 610
QY 541 TTCTCGAGTGTACAGGCGATACCGCTCCAGCTGAGGAGCTCAACCTCGGGTCAATC 600
Db 611 TTCTCGAGTGTACAGGCGATACCGCTCCAGCTGAGGAGCTCAACCTCGGGTCAATC 670
QY 601 TCGAATATCAGAGCTCTTCTCGATTCCTGAGATTCCTGAGGCTTCTGAGGATGAAACAGGC 660
Db 671 TCGAATATCAGAGCTCTTCTCGATTCCTGAGATTCCTGAGGCTTCTGAGGATGAAACAGGC 730
QY 661 ATAGCGATATCTGTGGCGCAGTGGCTCTGCGGCGCAAGTGTGTTGSAACGTACATAACT 720
Db 731 ATAGCGATATCTGTGGCGCAGTGGCTCTGCGGCGCAAGTGTGTTGSAACGTACATAACT 790
QY 721 TTGGACAAGACCTGGAAGGGAGTGACACCTCGGCTCGCTGGAGCTGGAGAACTGGCC 780
Db 791 TTGGACAAGACCTGGAAGGGAGTGACACCTCGGCTCGCTGGAGCTGGAGAACTGGCC 850
QY 781 GAGCTGTGCGGTGAGTGGCTGTTGTGGAGGTGCTGAGGCTCCCAACCAAGCAGCTG 840
Db 851 GAGCTGTGCGGTGAGTGGCTGTTGTGGAGGTGCTGAGGCTCCCAACCAAGCAGCTG 910
QY 841 CTGCTCTGTGAGATGGCTGCAATGAGAACTGGGCAAGTCTGTGTGGCCAAAGTCAAA 900
Db 911 CTGCTCTGTGAGATGGCTGCAATGAGAACTGGGCAAGTCTGTGTGGCCAAAGTCAAA 970
QY 901 ATTCCGGAAGGACCATCTTAAACATGAGATGCTACCGTGAAGGTGGGTGAGCCAAA 960
Db 971 ATTCCGGAAGGACCATCTTAAACATGAGATGCTACCGTGAAGGTGGGTGAGCCAAA 1030
QY 961 GCCTATCTCTGGAAGACATCTTAAATCTAGTGGGCAAGGTCTCTGCTCACTGTTGAA 1020
Db 1031 GCCTATCTCTGGAAGACATCTTAAATCTAGTGGGCAAGGTCTCTGCTCACTGTTGAA 1090
QY 1021 GAGGATGACACCATCATGGAAGATTTGGTAGATAATCATGGCAAAAAAATCAAGTCTTAA 1080
Db 1091 GAGGATGACACCATCATGGAAGATTTGGTAGATAATCATGGCAAAAAAATCAAGTCTTAA 1150

RESULT 9
AAI60928
ID AAI60928 standard; cDNA; 1243 BP.
XX
AC AAI60928;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4917.
XX
KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WC200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US014263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
(HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41172.
XX
Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.
Claim 1; SEQ ID NO 4917; 10078pp; English.
XX
The invention relates to human nucleic acids (AAI57798-AAI61369) and the
encoded polypeptides (AAM38642-AAM42213) with nontropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 1243 BP; 298 A; 291 C; 385 G; 269 T; 0 U; 0 Other;

Query Match 83.2%; Score 898.2; DB 4; Length 1243;
Best Local Similarity 88.6%; Pred. No. 4.5e-260;
Matches 1041; Conservative 0; Mismatches 3; Indels 131; Gaps 1;
QY 1 ATGCGCTGGAGCTGGAGCTGTGTCCCGGGCGCTGGTGGTGGCGGCAACACCCGCTGCTC 60
|||||
|||||

Db 69 ATGCCGCTGGAGCTGGAGCTGTCTCCGGGCGCTGGGTGGGGGCAACACCCGCTCTTC 128
QY 61 ATCAATGCCGATCGGCCAGAACCAACAGGGCGACCTGGACGTAGCCAAAGCGCATGATC 120
Db 129 ATCAATGCCGATCGGCCAGAACCAACAGGGCGACCTGGATGTAGCCAAAGCGCATGATC 188
QY 121 CCATGGCCCAAGAGTGTGGGCTGATCTGCCAGTTCACAGAGAGTGTAGTAATTC 180
Db 189 CCGATGGCCCAAGAGTGTGGGCTGATCTGCCAGTTCACAGAGAGTGTAGTAATTC 248
QY 181 AAGTTTAATCGAAAGCTTTGGAGAGCCATACACCTCGAAGCATTTCCTGGGGAGAGCG 240
Db 249 AAGTTTAATCGAAAGCTTTGGAGAGCCATACACCTCGAAGCATTTCCTGGGGAGAGCG 308
QY 241 TAGGGGAGACAAAGACATCTGGAGTTTCAGGCATTCACAGTACAGGAGCTGCAGAG 300
Db 309 TAGGGGAGACAAAGACATCTGGAGTTTCAGGCATTCACAGTACAGGAGCTGCAGAG 368
QY 301 TAGCCCGAGAGCTTGGGATCTTCTTCACTGCTCTGSCATGATGAGATGGCAGTTGAA 360
Db 369 TAGCCCGAGAGCTTGGGATCTTCTTCACTGCTCTGSCATGATGAGATGGCAGTTGAA 428
QY 361 TTCTGCACTAACTGAATCTTCATTTTCAAGTTTSGATCTGGAGACACTAATAATTTT 420
Db 429 TTCTGCACTAACTGAATCTTCATTTTCAAGTTTSGATCTGGAGACACTAATAATTTT 468
QY 421 CTTTATCTGAAAAGACAGCAAAA 446
Db 489 CTTTATCTGAAAAGACAGCAAAAAGGGCTGGCACAGTGTTTTAAGAGATGTCGTG 548
QY 447 ----- 446
Db 549 GAGTTCAGTTGATGATGATGAGACAGCAGTGGATGTGTGGGAGAGTCAGAACCTCTA 608
QY 447 -----AGTCCGCCAATGGTGATCTCA 469
Db 609 AAGAGAAAGTCTGATGTGTGGTCTCTGATTACTCAGTCCGCCAATGGTGATCTCA 668
QY 470 GTGGATGAGTCAATGGACACCAAGCAAGTTTATCAGATCGTGAAGCCCTCAACC 529
Db 669 GTGGATGAGTCAATGGACACCAAGCAAGTTTATCAGATCGTGAAGCCCTCAACC 728
QY 530 CCAACTCTCTCTTCTGCACTGATGATGATGATGATGATGATGATGATGATGATGATG 589
Db 729 CCAACTCTCTCTTCTGCACTGATGATGATGATGATGATGATGATGATGATGATGATG 788
QY 590 TCGGGGTATCTCGGAATATCAGAGCTCTTCTGCACTGATGATGATGATGATGATGATG 649
Db 789 TCGGGGTATCTCGGAATATCAGAGCTCTTCTGCACTGATGATGATGATGATGATGATG 848
QY 650 ATGAAACAGGATAGCGATCTGTGGCGCAGTGGCTCTGGGGCCAGAGGTGTGGGAC 709
Db 849 ATGAAACAGGATAGCGATCTGTGGCGCAGTGGCTCTGGGGCCAGAGGTGTGGGAC 808
QY 710 GTCACTAACTTTGGAACAAGCTGGAAGGAGTGCACCTCGGCTCGCTGGAGCTGTG 769
Db 909 GTCACTAACTTTGGAACAAGCTGGAAGGAGTGCACCTCGGCTCGCTGGAGCTGTG 968
QY 770 GAGAACTGCCGAGCTGTGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 829
Db 965 GAGAACTGCCGAGCTGTGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 1028
QY 830 CCAAGCAGCTGCTGCCCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 889
Db 1029 CCAAGCAGCTGCTGCCCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1088
QY 890 CCAAGTGAATTTCCGGAAGGACCAATTTCAACATGACATGCTCACTGAGGCTGG 949
Db 1089 CCAAGTGAATTTCCGGAAGGACCAATTTCAACATGACATGCTCACTGAGGCTGG 1148
QY 950 GTGAGCCCAAGCTTATCTCTGGAAGCACTTTTAATCTAGTGGGCAAGAGGCTCTGG 1009
Db 1149 GTGAGCCCAAGCTTATCTCTGGAAGCACTTTTAATCTAGTGGGCAAGAGGCTCTGG 1208

QY 1010 TCACTGTTGAAGAGATGACACCATCATCATGGAAGAA 1044
Db 1239 TCACTGTTGAAGAGATGACACCATCATCATGGAAGAA 1243
RESULT 10
AAS33141/c
ID AAS33141 standard; cDNA; 1238 BP.
XX
AC AAS33141;
XX
DT 04-DEC-2001 (first entry)
XX
DE DNA encoding human secreted protein, Seq ID No 100.
XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cytoprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing; ss.
XX
OS Homo sapiens.
PN NC200155326-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001347.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234957P.
 PR 25-SEP-2000; 2000US-0234958P.
 PR 25-SEP-2000; 2000US-0234959P.
 PR 25-SEP-2000; 2000US-0234960P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 27-SEP-2000; 2000US-0236327P.
 PR 27-SEP-2000; 2000US-0236328P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239936P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241803P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246539P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.

PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-451931/48.
 P-PSDB; AAU20432.

New nucleic acids and polypeptides, useful for diagnosing, preventing or
 treating medical conditions.
 Claim 1; SEQ ID NO 100; 753pp; English.

The invention relates to novel isolated nucleic acid molecules (I)
 encoding human secreted proteins (II). (I) and (II) are used to prevent,
 treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
 goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
 the prevention, treatment and diagnosis of diseases associated with
 inappropriate expression of secreted proteins. (I) and complementary
 sequences may also be used as DNA probes in diagnostic assays (e.g.
 polymerase chain reactions (PCR)) to detect and quantitate the presence
 of similar nucleic acid sequences in samples, and so which patients may
 be in need of restorative therapy. (II) may also be used as antigens in
 the production of antibodies and in assays to identify modulators
 (agonists and antagonists) of the expression and activity of the secreted
 proteins. The anti-(II) antibodies and antagonists may also be used to
 down regulate expression and activity of (II). The anti-(II) antibodies
 may also be used as diagnostic agents for detecting the presence of (II)
 in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
 disorders include for example: immune/autoimmune diseases (e.g. HIV
 (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 melanomas, neoplasms of the breast or liver, Sezary syndrome and
 Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 Parkinson's disease and Charcot-Marie-Tooth disease), cardiac-/
 cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and
 thrombosis), infections caused by bacteria, viruses and fungi and ocular
 disorders (e.g. corneal infections). (I) and (II), agonists, antagonists
 and antibodies can also be used to promote wound healing, maintain organs
 and before transplantation, and support cell culture of primary tissues.

Query Match 67.8%; Score 732; DB 4; Length 1238;
 Best Local Similarity 99.7%; Pred. No. 6.5e-210;
 Matches 732; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 347 AGATGGCAGTTGAATTCCTGCGAAGACTGAATGTTCCATTTTCAAGTTCGATCGAG 406
 |||||
 Db 799 AGATGGCAGTTGAATTCCTGCGAAGACTGAATGTTCCATTTTCAAGTTCGATCGAG 740
 |||||
 QY 407 ACACTATATTTTCTTCTGAAAGACAGCAAAAGGTCGCCATGTGATCT 466
 |||||

Db 491 CCTTATCTGAAAGACAGNCAAAAGAGTGCCTCAATGGTGTCTCCAGTGGATCGAG 550
 QY 481 TCAATGACACCATGAGCAAGTTTATCATGATGTGAGGCCCTCAACCCCAACTTTCG 540
 Db 551 TCAATGACACCATGAGCAAGTTTATCATGATGTGAGGCCCTCAACCCCAACTTTCG 610
 QY 541 TTCTTCAGTGTACCAAGCGCATACCGCTCCAGCTGAGGACGTCACACCTGGGGTATC 600
 Db 611 TTCTTCAGTGTACCAAGCGCATACCGCTCCAGCTGAGGACGTCACACCTGGGGTATC 670
 QY 601 TCGGATATCAGAAAGCTTTTCTTGACATCCCATAGGATTTCTGGGCATGAACAGGC 660
 Db 671 TCGGATATCAGAAAGCTTTTCTTGACATCCCATAGGATTTCTGGGCATGAACAGGC 730
 QY 661 ATAGCGATATCTGTGGCCGAGTGGC-TCTGGGGGCCAAGTG-TGAAAGTGCATATAC 719
 Db 731 ATAGCGATATCTGTGGCCGAGTGGC-TCTGGGGGCCAAGTG-TGAAAGTGCATATAC 789
 QY 720 TTTGGACAGACCTGGAAGGGAG 743
 Db 790 TTTGGACAGACCTGGAAGGGAG 813

RESULT 12

AAH11876/C

ID AAH11876 standard; cDNA; 514 BP.

XX AAH11876;

XX AC

XX DT

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (3'-primer) SEQ ID NO:8711.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX OS

XX EP1074617-A2.

XX PN

XX 07-FEB-2001.

XX PD

XX 28-JUL-2000; 2000EP-00116126.

XX PF

XX 29-JUL-1999; 99JP-00248036.

XX PR

XX 27-AUG-1999; 99JP-00300253.

XX PR

XX 11-JAN-2000; 2000JP-00118776.

XX PR

XX 02-MAY-2000; 2000JP-00183767.

XX PR

XX 09-JUN-2000; 2000JP-00241899.

XX XX

XX (HELI-) HELIX RES INST.

XX PA

XX Ora T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX PI

XX WPI; 2001-3-8749/34.

XX XX

XX Claim 3; SEQ ID NO 8711; 2537pp + Sequence Listing; English.

XX PS

XX The present invention describes primer sets for synthesizing 5602 full-

XX CC

XX length cDNAs defined in the specification. Where a primer set comprises:

XX CC (a) an oligo-dT primer and an oligonucleotide complementary to the

XX CC complementary strand of a polynucleotide which comprises one of the 5602

XX CC nucleotide sequences defined in the specification, where the

XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX CC of an oligonucleotide comprising a sequence complementary to the

XX CC complementary strand of a polynucleotide which comprises a 5'-end

XX CC sequence and an oligonucleotide comprising a sequence complementary to a

XX CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

XX SQ Sequence 514 BP; 112 A; 149 C; 113 G; 133 T; 0 U; 7 Other;

Query Match 44.8%; Score 483.8; DB 4; Length 514;
 Best Local Similarity 98.2%; Pred. No. 3.3e-135;
 Matches 485; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 587 ACCTGCGGTCATCTCGGATATCAGAAAGCTTTCTCCCTGACATCCCATAGGTTATCTG 546
 Db 514 ACNTGNGGGTNTNTNGGAANATCAGAAAGCTTTCTCCCTGACATCCCATAGGTTATCTG 455
 QY 647 GGCATGAAACAGGATAGCGATATCTGTGCGCGAGTGGCTCTGGGGGCCAAGGTGTTGG 706
 Db 454 GGCATGAAACAGGATAGCGATATCTGTGCGCGAGTGGCTCTGGGGGCCAAGGTGTTGG 395
 QY 707 AACGTCAATAAATTGGAACAAGACCTGGAAGGGAGTGACCACTCGGCGCTCGCTGGAGC 766
 Db 394 AACGTCAATAAATTGGAACAAGACCTGGAAGGGAGTGACCACTCGGCGCTCGCTGGAGC 335
 QY 767 CTGGAGAACTGGCGGAGCTGTGGGTGAGTGGCTCTGTGGAGCGTCCCTGGGCTCCC 826
 Db 334 CTGGAGAACTGGCGGAGCTGTGGGTGAGTGGCTCTGTGGAGCGTCCCTGGGCTCCC 275
 QY 827 CAACCAAGCAGCTGCTGCCCTGTGAGATGGCTTGCATGAGAAGCTGGGCAAGTCTGTGG 886
 Db 274 CAACCAAGCAGCTGCTGCCCTGTGAGATGGCTTGCATGAGAAGCTGGGCAAGTCTGTGG 215
 QY 987 TGGCCAAAGTGAAATTCGGGAAGGCCCACTTCTAACAATGGAATGTCACCGTGAAGG 946
 Db 214 TGGCCAAAGTGAAATTCGGGAAGGCCCACTTCTAACAATGGAATGTCACCGTGAAGG 155
 QY 947 TGGTGTAGCCCAAGCCCTATCTCTGAGACATCTTAACTAGTGGGCAAGAGGTCC 1006
 Db 154 TGGTGTAGCCCAAGCCCTATCTCTGAGACATCTTAACTAGTGGGCAAGAGGTCC 95
 QY 1007 TGGTCACTGTGGAAGAGGATGACACCATCATGGAGAATTTGGTAGATAATCATGGCAAAA 1066
 Db 94 TGGTCACTGTGGAAGAGGATGACACCATCATGGAGAATTTGGTAGATAATCATGGCAAAA 35

QY 1067 AAATCAAGTCTTAA 1080

Db 34 AAATCAAGTCTTAA 21

RESULT 13

AAS91837

ID AAS91837 standard; cDNA; 1417 BP.

XX AC

XX AAS91837;

XX XX

XX 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #27641.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX XX

XX WO200175067-A2.

XX PN

XX XX

PD 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US008631.
 PF 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649157.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 XX P-PSDB; AB327650.
 DR New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 27641; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1417 BP; 353 A; 336 C; 409 G; 319 T; 0 U; 0 Other;
 Query Match 39.5%; Score 426.4; DB 5; Length 1417;
 Best Local Similarity 89.4%; Pred. No. 1.2e-117;
 Matches 588; Conservative 0; Mismatches 46; Indels 24; Gaps 11;
 QY 447 AGGTGCGCCCAATGTTGATCTCCAGTGGGATGCAATGACACCATGACCAAGCAAGTTTA 506
 DB 427 AGGTGCGCCCAATGTTGATCTCCAGTGGGATGCAATGACACCATGACCAAGTTAT 486
 QY 507 TCAGATGCTGAAGCCCTCAACCCCAATCTCTGCTTC-TTGGAGTGTACCGCATACC 565
 DB 487 TCAAAATGCTGAAGCCCTCAACCCCAATCTCTGCTTC-TTGGAGTGTACCGCATACC 546
 QY 566 CGCTCCAGCTGAGGACCTCAACC--TGGGCTCATCTCGGAATATCAGAGCTCTTCC 623
 DB 547 CGCTCCAGCTGAGGACCTCAACCCTCGCGGTCACTCGGAATATCAGAGCTCTTCC 606
 QY 624 TGACATTTCCC---ATAGAGGTATCTTGGGCATGAACACAGGCATAGGATATCTGT---GGC 677
 DB 607 TGACATTTCCCATTAGGAGTATCTTGGGCATGAACACAGGCATAGGATATCTTGGCC 666
 QY 678 CGCAGTGGCTTGGGG--CCAGGTGTGGAA-----CGTACATAACTTTGGCAAGAC 731
 DB 667 CCCAGTGGCTTGGGGGCCCAAGGTGTGGAACTCTCAATAACTTTGGCAAGAC 726
 QY 732 CTGGAAGGGGAGTG---ACCACTCGGCTCGCT--GGAGCTGGAGAACTGGCGG-AGC 784
 DB 727 CGGGAGAGGGGAGTGTGACCCACTCGGCTCGCTTGGAGCCCTTGGAGAACTGGCGGAGC 786

QY 785 TGGTGGGTGAGTGGCTCTTGTGGAGCGTGCCTGGGTCTCCCAACC--AAGCAGCTGCTG 843
 DB 787 TGGTGGGTGAGTGGCTCTTGTGGAGCGTGCCTGGGTCTCCCAACCCTAGCAGCTGCTG 846
 QY 844 CCTGTGAGATGGCTGCAA--TGAGAACTGGGCAAGTCTGTGGTGGCCAAAGTGAATAAT 902
 DB 847 CCTGTGAGATGGCTGCAA--CTGGGCAAGTCTGTGGTGGCCAAAGTGAATAAT 906
 QY 903 TCCGGAAGGCCACCATTTCTAACTGACATGCTCACTCCCTGAGAGTGGGTGAGCCCAAGC 962
 DB 907 TCCGGAAGGCCACCATTTCTAACTGACATGCTCACTCCCTGAGAGTGGGTGAGCCCAAGC 966
 QY 963 CTATCTCTCTGAAGACATCTTTAACTAGTGGGCAAGAGGTCTCTGCTCACTGTTGAGA 1022
 DB 967 CTATCTCTCTGAAGACATCTTTAACTAGTGGGCAAGAGGTCTCTGCTCACTGTTGAGA 1026
 QY 1023 GGATGACACCATCATGGAAGAAATGGTGTAGATTAATCATGGCAAAAATAATCAAGTCTTAA 1080
 DB 1027 GGATGACACCATCATGGAAGAAATGGTGTAGATTAATCATGGCAAAAATAATCAAGTCTTAA 1084
 RESULT 14
 AEN95016/c
 ID AEN95016 standard; DNA; 464 BP.
 XX AEN95016;
 AC AEN95016;
 XX DT 13-AUG-2002 (first entry)
 XX DE Gene #1514 used to diagnose liver cancer.
 XX KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 XX metastatic liver tumor; cytostatic; expression profile; disease state;
 XX disease progression; drug toxicity; drug efficacy; drug metabolism.
 XX OS Homo sapiens.
 XX PN WC200229103-A2.
 XX PD 11-APR-2002.
 XX PF 02-OCT-2001; 2001WO-US030589.
 XX PR 02-OCT-2000; 2000US-0237054P.
 XX (GENE-) GENE LOGIC INC.
 XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 WFI; 2002-426119/45.
 Diagnosing and detecting the progression of liver cancer, hepatocellular
 carcinoma or metastatic liver tumor in a patient, involves detecting the
 level of expression of two or more genes in a liver tissue sample.
 Claim 1; SEQ ID NO 1514; 298pp; English.
 The invention relates to a novel method for diagnosing and detecting the
 progression of liver cancer, hepatocellular carcinoma or metastatic liver
 tumor in a patient, and differentiating metastatic liver cancer from of
 hepatocellular carcinoma in a patient, involving detecting the level of
 expression of two or more genes represented in ABN93503-ABN97455 in a
 tissue sample. The method of the invention has hepatotropic, and
 cytostatic activity. The method is useful for diagnosing and detecting
 the progression of liver cancer, hepatocellular carcinoma and metastatic
 liver carcinoma in a patient. The method is useful for identifying
 expression profiles which serve as useful diagnostic markers as well as
 markers that can be used to monitor disease states, disease progression,
 drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
 for this patent did not form part of the printed specification, but was
 obtained in electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences

[illegible]

GenCore version: 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:33 ; Search time 3122.76 Seconds
(without alignments)
10327.779 Million cell updates/sec

Title: US-09-930-440B-5
Perfect score: 1080
Sequence: 1 atgcgcgtgagctggagct.....gcaaaaaaatcaatctcttaa 1080

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 1493130276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba.*

2: em_estim.*

3: em_estim.*

4: em_estim.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hic.*

9: gb_esti.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estim.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vri.*

28: gb_gssi.*

29: gb_gss2.*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1076.8	99.7	1080	29	AY414837 Homo sapi
C 2	952.6	88.2	984	9	AL578658 AL578658
C 3	949.6	87.9	201	13	BX384733 BX384733
C 4	917.8	85.0	990	9	AL578579 AL578579

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	883.2	81.8	1080	29	AY414839	Mus muscu
6	883.2	81.8	1906	11	AK076290	Mus muscu
7	870.6	80.6	2109	11	AK088912	Mus muscu
8	856.6	79.3	925	12	B1517350	603041586
9	856.4	79.3	987	9	AL556387	AL556387
10	843	78.1	1005	9	AL556278	AL556278
11	840	77.8	1201	9	AL515470	AL515470
12	830	76.9	989	9	AL526241	AL526241
13	823.4	76.2	889	13	BU187658	AGENCOURT
14	803.2	74.4	872	13	BQ432284	AGENCOURT
15	796.4	73.7	1031	13	BQ064521	AGENCOURT
16	795.8	73.7	842	9	AL522851	AL522851
17	792.2	73.4	1049	12	BM541313	AGENCOURT
18	789.4	73.1	948	29	AY414838	Pan trogl
19	786	72.8	1046	9	AL526281	AL526281
20	784.4	72.6	890	13	BU181010	AGENCOURT
21	775	71.8	979	12	BG681730	AGENCOURT
22	772	71.5	905	13	BQ893527	AGENCOURT
23	764.4	70.8	1097	12	BM806099	AGENCOURT
24	757.4	70.1	878	9	AL522852	AL522852
25	754.4	69.9	878	12	B1824205	AL544111
26	749.2	69.4	1070	9	AL544111	AL544111
27	737.6	68.3	1050	12	BM563430	AGENCOURT
28	736.2	68.2	773	12	B1517963	603041586
29	735.4	68.1	845	13	BU528624	AGENCOURT
30	726.2	67.2	905	13	BQ649507	AGENCOURT
31	722.2	66.9	883	12	B1258475	602972467
32	721.6	66.8	777	12	BG568099	602586980
33	717.4	66.4	978	12	BG114903	602315105
34	710.6	65.8	919	13	BU195376	AGENCOURT
35	710.2	65.8	813	9	AU132828	AU132828
36	703.2	65.1	1031	9	AL544197	AL544197
37	695	64.4	706	12	BG746292	602703522
38	690	63.9	690	12	BG746329	602703579
39	689.4	63.8	835	10	BE614630	601504568
40	687.8	63.7	948	10	BE793764	601589915
41	686.4	63.6	769	13	BU618375	UI-H-PH1-
42	682.2	63.2	779	14	CA417480	UI-H-PE0-
43	680.4	63.0	931	12	B1525002	602924077
44	678.4	62.8	844	13	BU167760	AGENCOURT
45	675.2	62.5	877	12	B1905036	603169320

ALIGNMENTS

RESULT 1	AY414837	Homo sapiens NANS gene, VIRTUAL TRANSCR-PT, partial sequence,	1080 bp	DNA	linear	GSS 17-DBC-2003
LOCUS	AY414837	Genomic survey sequence.				
DEFINITION	AY414837.1	GI:39770796				
ACCESSION	AY414837	GSS.				
VERSION	AY414837.1	GI:39770796				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES

source Location/Qualifiers

gene 1..1080

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

<..>.1080

/gene="NANS"

/locus_tag="HCM5357"

ORIGIN

Query Match 99.7%; Score 1076.8; DB 29; Length 1080;

Best Local Similarity 99.8%; Pred. No. 3.4e-268;

Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCCGCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGGCAACACCGGTGCTTC 60

Db 1 ATCCCGCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGGCAACACCGGTGCTTC 60

QY 61 ATCATTTGCCAGATGCGCCAGACACACACAGGCGAGCTGACGTAGCCAGCGCATGATC 120

Db 61 ATCATTTGCCAGATGCGCCAGACACACACAGGCGAGCTGACGTAGCCAGCGCATGATC 120

QY 121 CGCATGGCCAAAGAGTGTGGGGCTGATTGTGCCAAGTTCCAGAGAGTGTAGATTC 180

Db 121 CGCATGGCCAAAGAGTGTGGGGCTGATTGTGCTAAGTTCCAGAGAGTGTAGATTC 180

QY 181 AAGTTTAATCGGAAGSCCTTGGAGGCGCCATACACCTCGAAGCATTCCTGGGGAGACG 240

Db 181 AAGTTTAATCGGAAGSCCTTGGAGGCGCCATACACCTCGAAGCATTCCTGGGGAGACG 240

QY 241 TAGCGGGAGACACAAAGACATCTGGAGTTTACGCATGACACAGTACAGGAGCTGCAGG 300

Db 241 TAGCGGGAGACACAAAGACATCTGGAGTTTACGCATGACACAGTACAGGAGCTGCAGG 300

QY 301 TAGCCGAGAGGTGTGGATCTTCTTCACTGCTCTGGCATGATGATGATGGAGTTGAA 360

Db 301 TAGCCGAGAGGTGTGGATCTTCTTCACTGCTCTGGCATGATGATGATGGAGTTGAA 360

QY 361 TTCTTCATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420

Db 361 TTCTTCATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420

QY 421 CCTTATCTGAAAGACAGCCAAAGAGTGTGGGGCTGATGATGATGATGATGATGATGAT 480

Db 421 CCTTATCTGAAAGACAGCCAAAGAGTGTGGGGCTGATGATGATGATGATGATGATGAT 480

QY 481 TCAATGGACACCATGAGCAAGTTTATCAGATCGTGAAGCCCTTCAACCCCAACTTCTGC 540

Db 481 TCAATGGACACCATGAGCAAGTTTATCAGATCGTGAAGCCCTTCAACCCCAACTTCTGC 540

QY 541 TTCTTCAGTGTACAGCGCATACCGCTCCAGCTGAGGAGCTGAACCTGGGGTCAATC 600

Db 541 TTCTTCAGTGTACAGCGCATACCGCTCCAGCTGAGGAGCTGAACCTGGGGTCAATC 600

QY 601 TCGGAATATCAGAGCTCTTCTCTGATTCCTGATTCCTGATTCCTGATTCCTGATTCCT 660

Db 601 TCGGAATATCAGAGCTCTTCTCTGATTCCTGATTCCTGATTCCTGATTCCTGATTCCT 660

QY 661 ATAGCGATATCTGTGGCGCAGTGGCTCTGTGGGGCCAAAGTGTGGAACTTCATTAAT 720

Db 661 ATAGCGATATCTGTGGCGCAGTGGCTCTGTGGGGCCAAAGTGTGGAACTTCATTAAT 720

QY 721 TTGGAACAAAGCTTGGAGGGAGTGAACCACTGGGCTTGTGGAGCTTGGAGAACTGGCC 780

Db 721 TTGGAACAAAGCTTGGAGGGAGTGAACCACTGGGCTTGTGGAGCTTGGAGAACTGGCC 780

QY 781 GAGCTGTGGGTGAGTGGCTCTGTGGAGCTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 840

Db 781 GAGCTGTGGGTGAGTGGCTCTGTGGAGCTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 840

QY 841 CTGCCCTGTGAGTGGCTGCAATGAGAGCTGGCAAGTCTGTGTGGGCGCAAGTGA 900

Db 841 CTGCCCTGTGAGTGGCTGCAATGAGAGCTGGCAAGTCTGTGTGGGCGCAAGTGA 900

QY 901 ATTCCGGAAGGCACCAATTTCAACAATGGACATCTCACCGTGAAGTGGTGGAGCCAAA 960

Db 901 ATTCCGGAAGGCACCAATTTCAACAATGGACATCTCACCGTGAAGTGGTGGAGCCAAA 960

QY 961 GCCTATCTCTCCGCAAGACATCTTTAATCTAGTGGCAAGAGTCTCTGGTCACTGTGAA 1020

Db 961 GCCTATCTCTCTCCGCAAGACATCTTTAATCTAGTGGCAAGAGTCTCTGGTCACTGTGAA 1020

QY 1021 GAGGATGACACCATCATGGAAGATTTGATAGATTAATCATGCGCAAAAAAATCAAGTCTTAA 1080

Db 1021 GAGGATGACACCATCATGGAAGATTTGATAGATTAATCATGCGCAAAAAAATCAAGTCTTAA 1080

RESULT 2

AL578658/c

LOCUS

DEFINITION

AL578658 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens cDNA clone CS0DK004Y114 3-PRIME, mRNA sequence.

ACCESSION

AL578658

VERSION

AL578658.2

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 984)

AUTHORS

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

On Feb 16, 2001 this sequence version replaced gi:12942945.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5047.f for more information about this cluster, see

http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DK004BE07NP2&cluster=5047.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DK004BE07NP1.

FEATURES

Location/Qualifiers

1..984

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DK004Y114"

/cell_line="HELA CELLS COT 25-NORMALIZED"

/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"

/notes="First prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 88.2%; Score 952.6; DB 9; Length 984;

Best Local Similarity 98.4%; Pred. No. 5.7e-236;

Matches 966; Conservative 3; Mismatches 12; Indels 1; Gaps 1;

QY 99 GGAAGTAGCCACCGCATATCCGATGCGCATGCGCAAGAGTGTGGGCTGATTTGCCAAGTT 158

Db 984 GGAAGTAGCCACCGCATATCCGATGCGCATGCGCAAGAGTGTGGGCTGATTTGCCAAGTT 925

QY 159 CCAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 218

Db 924 CCAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 865

```
QY 219 GAAGCATTCCTGGGGAGAGCGTACGGGAGGACAAACGACATCTGGAGTTCAAGCATGA 278
Db 864 GAAGCATTCCTGGGGAGAGCGTACGGGAGGACAAACGACATCTGGAGTTCAAGCATGA 805
QY 279 CCAAGTACAGGAGCTGAGAGGAGTACGGGAGGAGTGGGATCTTCTCACTGCTCTGG 338
Db 804 CCAAGTACAGGAGCTGAGAGGAGTACGGGAGGAGTGGGATCTTCTCACTGCTCTGG 745
QY 339 CATGGATGAGATGCGAGTTGAATTCCTGCACTGAATGAATGCTCCATTTTCAAAGTTGG 398
Db 744 CATGGATGAGATGCGAGTTGAATTCCTGCACTGAATGCTCCATTTTCAAAGTTGG 685
QY 399 ATCTGGAGACATTAATTTTCTTATCTTGGAAAGACAGCCAAAAGAGTTCGCCCAT 458
Db 684 ATCTGGAGACATTAATTTTCTTATCTTGGAAAGACAGCCAAAAGAGTTCGCCCAT 625
QY 459 GGTGATCTCCAGTGGGATGCAATGAGGACACCATGAAGCAAGTTTATCAGATCGTGAA 518
Db 624 GGTGATCTCCAGTGGGATGCAATGAGGACACCATGAAGCAAGTTTATCAGATCGTGAA 565
QY 519 GCCCTCAACCCCAACTTCTGCTTCTTGCAGTGTACAGCGCATACCGCTCCAGCCCTGA 578
Db 564 GCCCTCAACCCCAACTTCTGCTTCTTGCAGTGTACAGCGCATACCGCTCCAGCCCTGA 505
QY 579 GAGCTCAACTGCGGCTCATCTCGGAATATCAGAGCTCTTCTGATATCCCATAGG 638
Db 504 GAGCTCAACTGCGGCTCATCTCGGAATATCAGAGCTCTTCTGATATCCCATAGG 446
QY 639 GTATTCGGGATCAAAACAGGACATCTGTGCGCCAGCTGTCTGGGGGCGCAA 698
Db 445 GTATTCGGGATCAAAACAGGACATCTGTGCGCCAGCTGTCTGGGGGCGCAA 386
QY 699 GGTGTTGGAAAGCTCACATACTTTGGACAAAGCTTGAAGGGAGTGACCACTCGGCCTC 758
Db 385 GGTGTTGGAAAGCTCACATACTTTGGACAAAGCTTGAAGGGAGTGACCACTCGGCCTC 326
QY 759 GCTGGAGCTGGAGAACTGGCGGAGCTGGTGGCTGAGTGGCTCTTGTGGAGGCTGCCCT 818
Db 325 GCTGGAGCTGGAGAACTGGCGGAGCTGGTGGCTGAGTGGCTCTTGTGGAGGCTGCCCT 266
QY 819 GGGCTCCCAACCAAGCGCTGCTGCCCTGTGAGATGCGCTGCATGAGAGCTGGGCAA 878
Db 265 GGGCTCCCAACCAAGCGCTGCTGCCCTGTGAGATGCGCTGCATGAGAGCTGGGCAA 206
QY 879 GTCTGTGGTGGCCAAAGTGAATTCGGAAGGACCACTTCTAACAAATGACATGCTCAC 938
Db 205 GTCTGTGGTGGCCAAAGTGAATTCGGAAGGACCACTTCTAACAAATGACATGCTCAC 146
QY 939 CGTGAAGTGGGTGAGCCCAAGCTATCTCTCTGAGAGATCTTTAATCTAGTGGGCAA 998
Db 145 CGTGAAGTGGGTGAGCCCAAGCTATCTCTCTGAGAGATCTTTAATCTAGTGGGCAA 86
QY 999 GAAGGCTCTGCTCACTGTTGAAGAGGATGACACCATCATGGAAGAAATGGGTAGATAATCA 1058
Db 85 GAAGGCTCTGCTCACTGTTGAAGAGGATGACACCATCATGGAAGAAATGGGTAGATAATCA 26
QY 1059 TGGCAAAAAAATCAAGTCTTAA 1080
Db 25 TGGCAAAAAAATCANNNTNYMAM 4
```

```
RESULT 3
BX384733/c
LOCUS
DEFINITION BX384733 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
Homo sapiens cDNA clone CS0DL001YG22 3-PRIME, mRNA sequence.
ACCESSION BX384733
VERSION BX384733.1 GI:30449274
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Séquencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5047.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DL001BD11NPI&cluster=5047.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DL001BD11NPI.

FEATURES
source

Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL001YG22"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 87.9%; Score 949.6; DB 13; Length 1201;
Best Local Similarity 97.1%; Pred. No. 3.8e-235;
Matches 1039; Conservative 11; Mismatches 10; Indels 10; Gaps 8;
QY 1 ATGCGCTGGAGCTGGAGCTGTGCTCCCGGGCGCTGGCTGGGGCGGCAACACCGCTGCTTC 60
Db 1060 ATGCGCTGGAGYKG--AGTKTGTCCSGGCGCTGG--TGGGCGGCAACACCGCTGCTTC 1005
QY 61 ATCATTCGGAGTCCGCGAGACCCAGACCCAGAGCGGACCTGGAGCTGACCAAGCGGATGTC 120
Db 1004 ATCA-TGS-CGAGATC-GCCAGAACCCCA-GGCGACTGGAGCTGATCCAGCGCATGATC 948
QY 121 CSCATGSCCAAGAGTGTGGGGCTGATGTCGCCAAGTTCCAGAGAGTGAGTGAATTC 180
Db 947 GCGATGGCCAAAGGAGTGTG-GGCTGATGTGCTAAGTTCCAGAGAGTGAGTGAATTC 889
QY 181 AAGTTTAATCGAAAGCCTTGGAGGCGCATACCTCGAGAGCATTCCTGGGGGAGAGCG 240
Db 888 AAGTTTAATCGAAAGCCTTGGAGGCGCATACCTCGAGAGCATTCCT--GGGGGAGAGCG 830
QY 241 TACGGGAGGACAAACGACATCTGGAGTTCAGCCATGACCACTACAGGAGCTGCAGAGG 300
Db 829 TACGGGAGGACAAACGACATCTGGAGTTCAGCCATGACCACTACAGGAGCTGCAGAGG 770
QY 301 TACGCCGAGGAGTTGGGATCTTCTTCACTGCCCTCTGGCATGGATGAGATGSCAGTTGAA 360
Db 769 TACGCCGAGGAGTTGGGATCTTCTTCACTGCCCTCTGGCATGGATGAGATGSCAGTTGAA 710
QY 361 TTCTGTGATGAATCTGAATGTTCCATTTTCAAAGTTGGATCTGGAGACACTAATAATTTT 420
Db 709 TTCTGTGATGAATCTGAATGTTCCATTTTCAAAGTTGGATCTGGAGACACTAATAATTTT 650
QY 421 CCTTATCTGGAAGAGACAGCCAAAAGGCTCGCCCATGCTGCTCCAGTGGGATGAG 480
Db 649 CCTTATCTGGAAGAGACAGCCAAAAGGCTCGCCCATGCTGCTCCAGTGGGATGAG 590
QY 481 TCAATGACACCATGAAGCAAGTTTATCAGATCTGMAAGCCCTCAACCCCAACTTCTGC 540
Db 589 TCAATGACACCATGAAGCAAGTTTATCAGATCTGMAAGCCCTCAACCCCAACTTCTGC 530

```

QY 541 TTCTTCAGTGTACCGGACACACCGCTCCAGCTGAGGACGTCACACCTGCGGCTATC 600
Db |||||||
QY 529 TTCTTCAGTGTACCGGACACACCGCTCCAGCTGAGGACGTCACACCTGCGGCTATC 471
Db |||||||
QY 601 TCGGAATATCAGAACTCTTCTCTGACATCCCATAGGTAATCTGGGCACTGAACAGGC 660
Db |||||||
QY 470 TCGGAATATCAGAACTCTTCTCTGACATCCCATAGGTAATCTGGGCACTGAACAGGC 411
Db |||||||
QY 661 ATAGCCATATCTGCGCCGAGTGGCTCTGGGGCCCAAGTGTGGAAGTCACATAACT 720
Db |||||||
QY 410 ATAGCCATATCTGCGCCGAGTGGCTCTGGGGCCCAAGTGTGGAAGTCACATAACT 351
Db |||||||
QY 721 TTGGCAAGACCTGGAAGGAGTGCACCACTCGGCTCGCTGAGAGCTGGAGAACTGGCC 780
Db |||||||
QY 350 TTGGCAAGACCTGGAAGGAGTGCACCACTCGGCTCGCTGAGAGCTGGAGAACTGGCC 291
Db |||||||
QY 781 GAGCTGGTGGTCACTGGGCTCTTGTGGAGCGTGGCTCGGCTGCCCAACCAAGCAGCTG 840
Db |||||||
QY 290 GAGCTGGTGGTCACTGGGCTCTTGTGGAGCGTGGCTCGGCTGCCCAACCAAGCAGCTG 231
Db |||||||
QY 841 CTGCCCTGTCAGATGCGCTGCATGAGAGCTGGGCACTGCTGTGGGCAAGTGAA 900
Db |||||||
QY 230 CTGCCCTGTCAGATGCGCTGCATGAGAGCTGGGCACTGCTGTGGGCAAGTGAA 171
Db |||||||
QY 901 ATTCCGGAAGGACCACTCTTAACAATGGAATCTCAACCGTGAAGTGGGTGAGCCCAA 960
Db |||||||
QY 170 ATTCCGGAAGGACCACTCTTAACAATGGAATCTCAACCGTGAAGTGGGTGAGCCCAA 111
Db |||||||
QY 961 GCTATCTCTGGAAGACATCTTTAATCTAGTGGGCAAGAGTCTCTGTGTCATGTTGAA 1020
Db |||||||
QY 110 GGTATCTCTCTGGAAGACATCTTTAATCTAGTGGGCAAGAGTCTCTGTGTCATGTTGAA 51
Db |||||||
QY 1021 GAGGATGACCACTATCTGGAAGATCTGTAAGTAACTCATGGCAAAAAAT 1070
Db |||||||
QY 50 GAGGATGACCACTATCTGGAAGATCTGTAAGTAACTCATGGCAAAAAAT 1
Db |||||||

```

RESULT 4

```

AL578579/c
LOCUS AL578579 Homo sapiens HELA CELLS COT 25-NORMALIZED EST 01-JUN-2003
DEFINITION cDNA clone CS0DK001YM22 3-PRIME, mRNA sequence.
ACCESSION AL578579
VERSION AL578579.1 GI:12942789
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5047.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK001BG11NP1&cluster=5047.f. Contact :
Feng Liang Email : fliang@life-tech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DK001BG11NP1.
Location/Qualifiers
1. .990
/organism="Homo sapiens"
/db_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK001YM22"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
FEATURES
source

```

```

/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 85.0%; Score 917.8; DB 9; Length 990;
Best Local Similarity 96.8%; Pred. No. 6.2e-227; Indels 3; Gaps 3;
Matches 958; Conservative 8; Mismatches 23;
QY 1 ATCCCGCTGAGCTGGAGCTGTGTCCCGGCGTGGTGGGCGGCGGCAACCCGTCGCTTC 60
Db |||||||
QY 990 ATCCCGCTGAGCTGGAGCTGTGTCCCGGCGTGGTGGGCGGCGGCAACCCGTCGCTTC 931
Db |||||||
QY 61 ATCAATTCGCGAGATCGCGCGAGAACACACAGGGCGACCTGGACGTAGCGAGCGCATATC 120
Db |||||||
QY 930 ATAAATTCGCGAGATCGCGCGAGAACACACAGGGCGACCTGGATGTAGCCAGCGCATATC 871
Db |||||||
QY 121 CGCATGGCCAAAGAGTGTGGGCTGATTGTGCCAAGTTCACAGAGAGTGAAGTAGAATTC 180
Db |||||||
QY 870 CGCATGGCCAAAGAGTGTGGGCTGATTGTGCCAAGTTCACAGAGAGTGAAGTAGAATTC 811
Db |||||||
QY 181 AAGCTTTAATCGGAAAGCCTTTGGAGAGGCCATACACCTCGAAGCATTCCTCGGGGAAACAG 240
Db |||||||
QY 810 AAGCTTTAATCGGAAAGCCTTTGGAGAGGCCATACACCTCGAAGCATTCCTCGGGGAAACAG 751
Db |||||||
QY 241 TACGGGGAGCACAAAGACATCTGGAGTTCAGGCATGACCACTACAGGAGCTGCACAGG 300
Db |||||||
QY 750 TACGGGGAGCACAAAGACATCTGGAGTTCAGGCATGACCACTACAGGAGCTGCACAGG 691
Db |||||||
QY 301 TACGGCGAGGAGTGGGATCTTTCACTGCTCTGGCATGATGATGATGATGATGATGATGAA 360
Db |||||||
QY 690 TACGGCGAGGAGTGGGATCTTTCACTGCTCTGGCATGATGATGATGATGATGATGATGAA 632
Db |||||||
QY 361 TTCTTCGATGAACCTGAATGTTCATTTTCAAGTGTGGATCTCGAGACATCAATAAATTT 420
Db |||||||
QY 631 TTCTTCGATGAACCTGAATGTTCATTTTCAAGTGTGGATCTCGAGACATCAATAAATTT 572
Db |||||||
QY 421 CCTTATCTGGMAAGACAGCCAAAGAGTGGCCCATGCTGATCTCCAGTGGGATCAG 480
Db |||||||
QY 571 CCTTATCTGGMAAGACAGCCAAAGAGTGGCCCATGCTGATCTCCAGTGGGATCAG 512
Db |||||||
QY 481 TCAATGGAACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTCTTCG 540
Db |||||||
QY 511 TCAATGGAACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTCTTCG 452
Db |||||||
QY 541 TTCTTCGATGAACCTGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTCTTCG 600
Db |||||||
QY 451 TTCTTCGATGAACCTGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTCTTCG 393
Db |||||||
QY 601 TTGGATATACAGAACTCTTTCTTGACATTCCTCATAGGTAATCTCGGCAATGAGACAGC 660
Db |||||||
QY 392 TTGGATATACAGAACTCTTTCTTGACATTCCTCATAGGTAATCTCGGCAATGAGACAGC 333
Db |||||||
QY 661 ATAGCGATATCTGTGCGCCGAGTGGCTCTGGGGGCGCAAGTGTGTGAACCTGCATAACT 720
Db |||||||
QY 332 ATAGCGATATCTGTGCGCCGAGTGGCTCTGGGGGCGCAAGTGTGTGAACCTGCATAACT 273
Db |||||||
QY 721 TTGGCAAGACCTGGAAGGAGTGCACCACTCGGCTCGCTGAGAGCTGGAGAACTGGCC 780
Db |||||||
QY 272 TTGGCAAGACCTGGAAGGAGTGCACCACTCGGCTCGCTGAGAGCTGGAGAACTGGCC 213
Db |||||||
QY 781 GAGCTGGTGGTCACTGGGCTCTTGTGGAGCGTGGCTCGGCTGCCCAACCAAGCAGCTG 840
Db |||||||
QY 212 GAGCTGGTGGTCACTGGGCTCTTGTGGAGCGTGGCTCGGCTGCCCAACCAAGCAGCTG 153
Db |||||||
QY 841 CTGCCCTGTCAGATGCGCTGCATGAGAGCTGGGCACTGCTGTGTGGGCAAGTGAA 900
Db |||||||
QY 152 CTGCCCTGTCAGATGCGCTGCATGAGAGCTGGGCACTGCTGTGTGGGCAAGTGAA 93
Db |||||||
QY 901 ATTCCGGAAGGACCACTCTTAACAATGGAATCTCAACCGTGAAGTGGGTGAGC-CCAA 959
Db |||||||

```



```

Db      92  ATCCGGAGGACCACTTCTMACAAATGAACATGCTCACCGTGAAGGTGGGTGAGCGCAA 33
Qy      960 AGCTATCTCTCTGAAGACATCTTTAATCTAG 991
Db      32  AGGCTATCTCTCTGAAGACATCTTTAATCTAG 1

RESULT 5
AY414839      1080 bp      DNA      linear      GSS 17-DEC-2003
DEFINITION   Mus musculus NANS gene, VIRTUAL TRANSCRIPT, partial sequence,
              genomic survey sequence.
ACCESSION    AY414839
VERSION      AY414839.1 GI:39770798
KEYWORDS     GSS.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE    1 (bases 1 to 1080)
AUTHORS      Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
              Todd,X.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
              Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
              Adams,M.D. and Cargill,M.
              Inferring nonneutral evolution from human-chimp-mouse orthologous
              gene trios
JOURNAL      Science 302 (5652), 1960-1963 (2003)
PUBMED      14671302
REFERENCE    2 (bases 1 to 1080)
AUTHORS      Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
              Todd,X.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
              Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
              Adams,M.D. and Cargill,M.
              Direct Submission
              Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
              Rockville, MD 20850, USA
              This sequence was made by sequencing genomic exons and ordering
              them based on alignment.
FEATURES     source      location/Qualifiers
              1..1080
              /organism="Mus musculus"
              /mol_type="genomic DNA"
              /db_xref="taxon:10090"
              <1..>1080
              /gene="NANS"
              /locus_tag="HCM5357"
ORIGIN
Query Match      81.8%; Score 883.2; DB 29; Length 1080;
Best Local Similarity 88.6%; Pred. No. 6.3e-218;
Matches 95; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy      1  ATCCCGCTGGAGCTGGAGCTGTGTCCGGGGCGCTGGGTGGCGGGCAACACCCGTGCTTC 60
Db      1  ATCCCGCTGGAACTGGAGCTGTGTCCGGGGCGCTGGGTGGGTGGAAAGCACCCGTGCTTC 60
Qy      61  ATCATTCGAGATCGGCGAAGACCAACAGGCGGACTGGACGTAGCAAGCGCATGATC 120
Db      61  ATCATTCGAGATCGGCGAAGACCAACAGGCGGACTGGACGTAGCAAGCGCATGATC 120
Qy      121  CGCATGCGCAAGGAGTGTGGGGCTGTGTCAGGTTCCAGAGAGTGGAGTGAATTC 180
Db      121  CGCATGCGCAAGGAGTGTGGGGCTGTGTCAGGTTCCAGAGAGTGGAGTGAATTC 180
Qy      181  AAGTTTAAATCGAAAGCTTGGAGAGGCGATACACCTCGAAGCATTCCTGGGGGAAGACG 240
Db      181  AAGTTTAAACCGAAGGCGCTGGAGAGACCATATATCTCGAAGCATTCATGGGGGAAGACG 240
Qy      241  TACGGGAGACAAACGACATCTCGAGTTACGCCATGACCAAGTACAGGAGCTCCAGAGG 300
Db      241  TATGGGAGACAAACGCGCATCTGGAATTCAGCCACGACCAAGTACAAAGGAGCTCCAGAGG 300
Qy      301  TACGCCGAGGAGGTGGGAATCTCTTCACTGCCTCTGCGATGGATGAGATGGCAGTTGAA 360

```


Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 204939374
 11042159
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, K., Kishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5
 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 1306)
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
 Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
 Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
 Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Komno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
 Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
 Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D.,
 Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
 Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
 Tejima, Y., Toyota, T., Yamamura, T., Yamazaki, I., Yasunishi, A.,
 Yoshida, K., Yoshino, N., Muramatsu, M. and Hayashizaki, Y.
 Direct Submision
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/.
 Location/Qualifiers
 1. 1906
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:4632418E04"
 /db_xref="MGI:1903268"
 /db_xref="taxon:10090"
 /clone="4632418E04"
 /tissue_type="skin"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="0 day neonate"
 32. 1111
 /note="unnamed protein product; putative"

sialic acid synthase [MGD|2149820, GB|NM_053179,
 evidence: BLASTN, 100%, match=1140]"
 /codon_start=1
 /protein_id="BAC36290.1"
 /translation="MPLELELCPRWVGKHPFCIIAETGQNEQGDIDVAKRMRTAK
 EGADCAKPKSELEFFKFKRKALEKDYPTKSGNGKTYGHEKRLHEESHOYKELQSYA
 QSIGIFFTASGDMEMAVEFLHNLNVPFFKVGSGDTNNPYLEKTAAGKHPMTLSSGMQ
 SMDTKQVYQIVKPLNPFCFLOCTISAYPILOPEDANLRVISEYKQLFPDIPISGHE
 TGAISVAALVAGKRLKRLHITLTKTKGSDHSDLSPEGLAEVLSVRLVRLALGSP
 TKOLLPCMACNEKLGKSVAVAKVPAGTTLTLDMLTVKVGEPKGPPEPDIFFNLAKKK
 VLVITIEDDTVMNEESVESHSKKIKA"
 ORIGIN
 Query Match 81.8%; Score 883.2; DB 11; Length 1306;
 Best Local Similarity 88.6%; Pred. No. 8.5e-218; Indels 0; Gaps 0;
 Matches 957; Conservative 0; Mismatches 123;
 1 ATGCCCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGGGCAACACCCGTGCTTC 60
 32 ATGCCCTGGAACTGGAGCTGTGTCCCGGCGCTGGGTGGCGGGGCAACACCCGTGCTTC 91
 61 ATCATTCGAGATCGGCCAGAACACACCGAGGCGACTGGACCTAGCAGTACAGACGATGATC 120
 92 ATCATTCGAGATCGGCCAGAACACACCGAGGCGACTGGACCTAGCAGTACAGACGATGATC 151
 121 CGCATGGCCAGAGGTGTGGGCTGATTCGCCAGTTCAGAGAGAGTGGAGTATTC 180
 152 CGCATGGCCAGAGGTGTGGGCTGATTCGCCAGTTCAGAGAGAGTGGAGTATTC 211
 181 AAGTTTAAATCGGAAAGCCCTGGAGAGCCATACACCTCGAAGCATTCCTGGGGGAAGACG 240
 212 AAGTTTAAATCGGAAAGCCCTGGAGAGCCATACATATCTTGGAGCATTCATGGGGGAAGACG 271
 241 TAGGGGAGACACAAAGCATCTGGAGTTTACGATACAGTACAGGAGCTGCAGAG 300
 272 TAGGGGAGACACAAAGCATCTGGAAATTCAGCCACAGTACAGGAGCTGCAGAG 331
 301 TAGCGCGAGGAGTTGGGATCTCTTCACTGCTCTGGCATGATGATGAGTGGAGTTGA 360
 332 TAGCGCGAGGAGTGGGATCTCTTCACTGCTCTGGCATGATGATGAGTGGAGTTGA 391
 361 TCCCTGCATGAATGAATGTTCCATTTTCAAGTTGGATCTGGAGACACTAATAATTTT 420
 392 TTTCTGCAGAACTGAATGTTTCCATTTTCAAGTTGGATCTGGAGACACTAATAATTTT 451
 421 CTTATCTGGAAAAGACAGCCAAAAGTGGCCCAATGGTGTATCTCCAGTGGAGTGCAG 480
 452 CCTTACCTGGAAAAGACAGCCAAAAGTGGCCCAATGGTGTATCTCCAGTGGAGTGCAG 511
 481 TCAATGGACACCATGAAGCAAGTTCATCAGATCGTGAAGCCCTCAACCCCAACTTCGC 540
 512 TCATGGACACCATGAAGCAAGTTCATCAGATCGTGAAGCCCTCAACCCCAACTTCGC 571
 541 TTTCTGCAGTACAGGCGCATACCGGTCGAGCTGAGGAGTGAAGTCAACTCGGGGGTATC 600
 572 TTTCTGCAGTACAGGCGCATACCGGTCGAGCTGAGGAGTGAAGTCAACTCGGGGGTATC 631
 601 TCGGATATCAGAGCTCTTTCCTGCATTCATCCATAGAGTATCTGGGATGAAACAGCG 660
 632 TCGGATATCAGAGCTCTTTCCTGCATTCATCCATAGAGTATCTGGGATGAAACAGCG 691
 661 ATAGCGATATCTGTGGCGCGAGTGGCTCTGGGGGGCAAGGTGTGGAAAGTCACTAAT 720
 692 ATCGGCATATCTGTGGCGCGAGTGGCTCTGGGGGGCAAGGTGTGGAAAGTCACTAAT 751
 721 TTGGAACAAGCTGGAAAGGAGTGAACCACTCGGCTCTGCTGGAGCTGGAGAACTGGCC 780
 752 TTGGAACAAGCTGGAAAGGAGTGAACCACTCGGCTCTGCTGGAGCTGGAGAACTGGCC 811
 781 GAGCTGGTGGCTGAGTGGCTGTGTGGAGTGGCTGGGCTCCGCAACCAAGAGCTG 840
 812 GAGCTGGTGGCTGAGTGGCTGTGTGGAGTGGCTGGGCTCCGCAACCAAGAGCTG 871

Qy 841 CTGCCCTGTGAGATGGCTGCAATGAGAAGCTGGGCAAGTCTGTGGTGGCCAAAGTGAAA 900
 Db 872 CTGCCCTGTGAGATGGCTGCAATGAGAAGCTGGGCAAGTCTGTGGTGGCCAAAGTGAAA 931
 Qy 901 ATTCCGGAAGCAGCATCTTAACATGACATGCTCACCGTGAGAGTGGTGGAGCCGAAA 960
 Db 932 ATCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 991
 Qy 961 GCCTATCTCTCTGAAGCATCTTAACTAGTGGCAAGAGGCTCTGGTCACTGTTCGA 1020
 Db 992 GCCTATCTCTCTGAAGCATCTTAACTAGTGGCAAGAGGCTCTGGTCACTGTTCGA 1051
 Qy 1021 GAGGATGACACCATCATGGAAGAATTTGGTAGATAATCATGCGCAAAAATAAATCAAGTCTTAA 1080
 Db 1052 GAAGATGACACCGTTCATGGAAGAATTCGTGGAAAGTTCACAGCAAGAAATCAAGGCTTAA 1111

RESULT 7
 AK088912
 LOCUS Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
 DEFINITION full-length enriched library, clone:E430030P04 product:sialic acid synthase, full insert sequence.
 ACCESSION AK088912.1 GI:26104987
 VERSION HTCC; CAP trapper.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 2
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,H., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499574
 PUBMED 11042159
 3
 Shibata,K., Itoh,H., Aizawa,K., Nagao,K., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Katsura,T., Tashiro,H., Itoh,M., Sumi,X., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Shikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 5
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 6
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,

Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL:http://genome.gsc.riken.go.jp/
 URL:http://fantom.gsc.riken.go.jp/
 Location/Qualifiers
 1. 2109
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NOD"
 /db_xref="FANTOM_DB:E430030P04"
 /db_xref="WGI:2427805"
 /db_xref="taxon:10090"
 /clone="E430030P04"
 /cell_type="thymic cells"
 /tissue_type="thymus"
 /csource_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="2 days neonate"
 41. 1121
 /note="putative sialic acid synthase (MGD|MG1:2149820, GB|NM_053179, evidence: BLASTN, 100%, match=1140)"

misc_feature
 41. 1121
 /note="putative sialic acid synthase (MGD|MG1:2149820, GB|NM_053179, evidence: BLASTN, 100%, match=1140)"

ORIGIN
 Query Match 80.6%; Score 870.6; DB 11; Length 2109;
 Best Local Similarity 88.4%; Pred. No. 1.7e-214;
 Matches 956; Conservative 0; Mismatches 124; Indels 1; Gaps 1;
 Qy 1 ATCCGCTGTGAGATGGAGCTGTGTCGCGCGCTGGTGGCGGCAACACCCGCTGCTC 60
 Db 41 ATCCGCTGTGAGATGGAGCTGTGTCGCGCGCTGGTGGCGGCAACACCCGCTGCTC 100
 Qy 61 ATCATTCGCGAGATCGGCCAGAACACCAGCGCGACCTGGACGTAGCCAGCGATGATC 120
 Db 101 ATCATTCGCGAGATCGGCCAGAACACCAGCGAGACATAGATGTGGCCAGCGATGATC 160
 Qy 121 CGCATGCCAAGAGTGTGGGCTGTGATGTGCCAAGTTCGAGAGAGTGGACCTAGATTC 180
 Db 161 CGCATGCCAAGAGTGTGGGCTGTGATGTGCCAAGTTCGAGAGAGTGGACCTAGATTC 220
 Qy 181 AAGTTTAAATCGGAAGCTTTGGAGAGCCATACACCTCGAAGCATTCCTGGGGGAAGAGC 240
 Db 221 AAGTTTAAATCGGAAGCTTTGGAGAGCCATACACCTCGAAGCATTCCTGGGGGAAGAGC 280
 Qy 241 TAGCGGAGCAGCAAAACACATCTGGAGTTCAGCCATGACCACTACAGGAGCTGCAGAGG 300
 Db 281 TAGCGGAGCAGCAAAACACATCTGGAGTTCAGCCATGACCACTACAGGAGCTGCAGAGG 340
 Qy 301 TAGCGGAGGAGGTGGGATCTCTTCACTGCTCTGCGATGGCATGGATGAGATGGCATGGAA 360


```

TITLE
JOURNAL
COMMENT
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:12398807.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5047.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK001BG11QPI&cluster=5047.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DK001BG11QPI.

FEATURES
source
1. .1005
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK001YM22"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match: 78.1%; Score 843; DB 9; Length 1005;
Best Local Similarity 95.5%; Pred. No. 1.7e-207;
Matches 846; Conservative 6; Mismatches 34; Indels 3; Gaps 0;

QY 1 ATGCCCTGGAGCTGGAGCTGTGTCCCGCGCGTGGTGGCGGCGACACCCGCTGCTTC 60
Db 120 ATGCCCTGGAGCTGGAGCTGTGTCCCGCGCGTGGTGGCGGCGACACCCGCTGCTTC 179

QY 61 ATCATTCGCGAGATCGCCAGAACCCACGAGCGGCGCTGACCTAGCAGAGCCATGATC 120
Db 180 ATCATTCGCGAGATCGCCAGAACCCACGAGCGGCGCTGACCTAGCAGAGCCATGATC 239

QY 121 CGCATGCCAGGAGGTGTGGGCTGATTTGCCAAGTTTCAGAGAGAGTGTAGATTC 180
Db 240 CGCATGCCAGGAGGTGTGGGCTGATTTGCCAAGTTTCAGAGAGAGTGTAGATTC 299

QY 181 AAGTTTAACTCGGAAGCTTGGAGAGGCGATACCTCGAAGATTTCTTGGGGAGAGC 240
Db 300 AANTNTAATCCGAANCCCTTNGAGAGCCATACCTCCAAACATTTCTGGGGAGANACG 359

QY 241 TAGCGGAGGACAAACGACATCTGGAGTTTCAGCCATGACACAGTACAGGAGCTSCAGG 300
Db 360 TAGCGGAGGACAAACGACATCTGGAGTTTCAGCCATGACACAGTACAGGAGNCTCCANAG 419

QY 301 TAGCGGAGGAGGTGGAGTCTTTCTTCACTGCTCTGGCATGATGAGATGGCAGTTGAA 360
Db 420 TAGCGGAGGAGGTGGAGTCTTTCTTCACTGCTCTGGCATGATGAGATGGCAGTCGAA 479

QY 361 TCCCTGCTGAATCTGAATGTTCAATTTTCAAAGTTGGATCTGGAGACATTAATTTT 420
Db 480 TCCCTGCTGAATCTGAATGTTCAATTTTCAAAGTTGGATCTGGAGACATTAATTTT 539

QY 421 CTTATCTGAAAGACAGACCAAAAGGTGGCCCAATGTGATCTCAGTGGGATGAC 480
Db 540 CTTATCTGAAAGACAGACCAAAAGGTGGCCCAATGTGATCTCAGTGGGATGAC 599

QY 481 TCAATGACACCATGAGCAAGTTTATCAGATCTGAGGCCCTCAACCCCAACTTCTGC 540
Db 600 TCAATGACACCATGAGCAAGTTTATCAGATCTGAGGCCCTCAACCCCAACTTCTGC 659

QY 541 TCTTTCAGTGTACCGGATACCCGCTCAGCTGAGACGTAACTTGGGGTCAATC 600
Db 660 TTCTTTCAGTGTACCGGATACCCGCTCAGCTGAGACGTAACTTGGGGTCAATC 719

```

```

QY 601 TCGGAATATCAGAGCTCTTTCTCTGATTCATCCATAGGCTATTTCTGGCATGAACAGGC 660
Db 720 TCGGAATATCAGAGCTCTTTCTCTGATTCATCCATAGGCTATTTCTGGCATGAACAGGC 779

QY 661 ATAGCCATATCTGTGGCCGAGTGGCTCTGGGCGCCCAAGCTGTGGAAACGTCAACATACT 720
Db 780 ATAGCCATATCTGTGGCCGAGTGGCTCTGGGCGCCCAAGCTGTGGAAACGTCAACATACT 839

QY 721 TTGGACRAGACCTCGAAGGAGGTGACCACTCGGCTCTCGCTGGAGCCCTGGAGAACTGGCC 780
Db 840 TTGGACRAGACCTCGAAGGAGGTGACCACTCGGCTCTCGCTGGAGCCCTGGAGAACTGGCC 899

QY 781 GAGCTGTGGGTGAGTGGCTCTGTGGGCGGCTGGCTGGGCTCCCAACCAAGCAGCTG 840
Db 900 GAGCTGTGGGTGAGTGGCTCTGTGGGCGGCTGGCTGGGCTCCCAACCAAGCAGCTG 959

QY 841 CTGCCCTGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGG 886
Db 960 CTGCCCTGTGAGAGGCTGCAATGAGAGCTGGGCAAGTCTGTGG 1005

RESULT 11
AL515470
LOCUS
DEFINITION
CS0DA002YI24 5-PRIME, mRNA sequence.
ACCESSION
AL515470
VERSION
AL515470.2 GI:30489137
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
COMMENT
On Feb 13, 2001 this sequence version replaced gi:12778963.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5047.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DA002B12QPI&cluster=5047.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DA002B12QPI.

FEATURES
source
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DA002YI24"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and EcoRV
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match: 77.8%; Score 840; DB 9; Length 1201;
Best Local Similarity 98.7%; Pred. No. 1.1e-206;
Matches 859; Conservative 7; Mismatches 2; Indels 2; Gaps 2;

QY 1 ATGCCCTGGAGCTGGAGCTGTTCCTCCGGCGCTGGGCGGCGGCAACCCGCTGCTTC 60
Db 103 ATGCCCTGGAGCTGGAGCTGTTCCTCCGGCGCTGGGCGGCGGCAACCCGCTGCTTC 162

```

QY 61 ATCATTTGCCGAGATCGGCCAGAACACACAGGCGCGACCTGGACGTAGCCAGAGCGCATGATC 120
 Db 163 ATCATTTGCCGAGATCGGCCAGAACACACAGGCGCGACCTGGATGTAGCCAGAGCGCATGATC 222
 QY 121 CGCATGGCCAGGAGCTGGGGCTGATTTGTGCGCAGTTCACAGAGAGTGAGCTAGATTC 180
 Db 223 CGCATGGCCAGGAGCTGGGGCTGATTTGTGCGCAGTTCACAGAGAGTGAGCTAGATTC 282
 QY 181 AAGTTTAAATCGAAAGCCTTGGAGAGGCGATACACTCGAACATTTCTTGGGGAGAGCG 240
 Db 283 AAGTTTAAATCGAAAGCCTTGGAGAGGCGATACACTCGAACATTTCTTGGGGAGAGCG 342
 QY 241 TACGGGGAGCAAAACGACATCTGGAGTTCACCCATGACAGTACAGGAGCTGCGAGG 300
 Db 343 TACGGGGAGCAAAACGACATCTGGAGTTCACCCATGACAGTACAGGAGCTGCGAGG 402
 QY 301 TACGCCGAGGAGTGGGATCTTTCTCACTGCTCTGGCATGGATGAGATGCGATTGAA 360
 Db 403 TACGCCGAGGAGTGGGATCTTTCTCACTGCTCTGGCATGGATGAGATGCGATTGAA 462
 QY 361 TTCTGTCATGAACCTGATGTTCCATTTTCAAAGTTGATCTGCGACACTAATATTTT 420
 Db 463 TTCTGTCATGAACCTGATGTTCCATTTTCAAAGTTGATCTGCGACACTAATATTTT 522
 QY 421 CTTATCTGGAAAGACAGCCAAAGAGTGCCTGACATGCTGATCTGAGTGGGATGCGAG 480
 Db 523 CTTATCTGGAAAGACAGCCAAAGAGTGCCTGACATGCTGATCTGAGTGGGATGCGAG 582
 QY 481 TCNATGGACACCATCAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCTGC 540
 Db 593 TCNATGGACACCATCAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCTGC 642
 QY 541 TTCTTGAGTGTACAGGCGCATACCGCTCCAGCTGAGGAGCTCAACCTGCGGGGTCTATC 600
 Db 643 TTCTTGAGTGTACAGGCGCATACCGCTCCAGCTGAGGAGCTCAACCTGCGGGGTCTATC 702
 QY 601 TCGGATATCAGAGCTCTTCTGACATTCCTAGGATTCCTGAGGATTCCTGGGATGAAACAGCG 660
 Db 703 TCGGATATCAGAGCTCTTCTGACATTCCTAGGATTCCTGAGGATTCCTGGGATGAAACAGCG 762
 QY 661 ATAGCGATATCTGCGCGCAGTGCCTCTGGGGCCCAAGTCTTGGAGAGCTCACATACT 720
 Db 763 ATAGCGATATCTGAGCGCAGTGCCTCTGGGGCCCAAGTCTTGGAGAGCTCACATACT 822
 QY 721 TTGACAAAGCCTGGAAGGGAGTGACCACTCGGCTCTGCGAGCCTGGAGAACTGGCC 780
 Db 823 TTGACAAAGCCTGGAAGGGAGTGACCACTCGGCTCTGCGAGCCTGGAGAACTGGCC 881
 QY 781 GAGCTGTCGCGTCACTGCTGAGAGGCTGCTGAGAGGCTGCTGAGGCTGCTGAGAGGCTG 840
 Db 882 GAGCTGTCGCGTCACTGCTGAGAGGCTGCTGAGAGGCTGCTGAGAGGCTGCTGAGAGGCTG 941
 QY 841 CTGCGCTGTGAGATGGCTGCAATGAGAG 870
 Db 942 CTGCGCTGTGAGAGG-CTGCAATGAGAG 970

RESULT 12
 Locus AL526241/c
 DEFINITION AL526241 Homo sapiens NEUROBLASTOMA COX 25-NORMALIZED Homo sapiens
 cDNA clone CS0DC016Y22 3-PRIME, mRNA sequence.
 AL526241
 ACCESSION AL526241.2 GI:31064102
 VERSION 1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 989)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization

JOURNAL
COMMENT

Unpublished (2001)
 On Feb 13, 2001 this sequence version replaced gi:12789734.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 5047.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DC016BGL1NP1&cluster=5047.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1680
 Faraday Avenue Genoscope sequence ID : CS0DC016BGL1NP1.

FEATURES
source

1..989
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DC016Y22"
 /tissue="Homo sapiens NEUROBLASTOMA COX 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 76.9%; Score 830; DB 9; Length 989;
 Best Local Similarity 93.0%; Pred. NO. 3.9e-204; Indels 6; Gaps 4;
 Matches 872; Conservative 34; Mismatches 26;
 QY 126 GCGCCAGGAGTGTGGGGCTGATGTGCCAAGTTCCAGAGAGTGAGCTAGAAATCAAGTT 185
 Db 94C KCCCAAGGAGTGTGGGCTGATGTG---CTAAGTCCAGAGAGTGAAGTCAAGTT 884
 QY 186 TAATCGGAAGCCTTGGAGAGGCCATACACCTCGAAGCATTTCTGGGGGAGAGCGTACGG 245
 Db 883 TAATCGGAAGCCTTGGAGAGGCCATACACCTCGAAGCATTTCTGGGGGAGAGCGTACGG 824
 QY 246 GGAGCACAAAACACATCTGGAGTTCAGCCATGACAGTACAGGAGCTGCAGAGGTACGC 305
 Db 823 GGAGCACAAAACACATCTGGAGTTCAGCCATGACAGTACAGGAGCTGCAGAGGTACGC 764
 QY 306 CGAGGAGTGTGGATCTTTCTTCACTGCTCTGGCATGGATGAGATGCGAGTTGAATTCCT 365
 Db 763 CGAGGAGTGTGGATCTTTCTTCACTGCTCTGGCATGGATGAGATGCGAGTTGAATTCCT 704
 QY 366 GCATCAACTGAATGTTCCATTTTCAAGTTGGATCTGGAGACATTAATATTTTCTTA 425
 Db 703 GCATCAACTGAATGTTCCATTTTCAAGTTGGATCTGGAGACATTAATATTTTCTTA 644
 QY 426 TCTGAAAGACAGCAAAAAGGTCGCCAATGCTGATCTCCAGTGGGATGCAGTCAAT 485
 Db 643 TTTGAAAGACAGCAAAAAGGTCGCCAATGCTGATCTCCAGTGGGATGCAGTCAAT 584
 QY 486 GGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCTGCTTCT 545
 Db 583 GGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCTGCTTCT 524
 QY 546 GCAGTGTACAGCATACCGCTCTCAGCTGAGAGAGCTCAACCTGCGGGTCACTTCGGA 605
 Db 523 GCAGTGTACAGCATACCGCTCTCAGCTGAGAGAGCTCAACCTGCGGGTCACTTCGGA 465
 QY 606 ATATCAGAGCTCTTTTCTGCAATTCCTATGGGTATTTCTGGGCATGAACAGGATAGC 665
 Db 464 ATATCAGAGCTCTTTTCTGCAATTCCTATGGGTATTTCTGGGCATGAACAGGATAGC 405
 QY 666 GATATCTGTGCGCGCAGTGGCTCTGGGGGCCAAGGTGTTGGAACCTCACATAACTTTTGA 725
 Db 404 GATATCTGTGCGCGCAGTGGCTCTGGGGGCCAAGGTGTTGGAACCTCACATAACTTTTGA 345
 QY 726 CAAGACCTGGAAGGGAGTG-ACCACCTCGGCTCGAGCCTCGAGAACTGGCCGAGC 784

```

Db      344  CAAAGACTGGAGGAGGAGGACCCSVCTCSCHGAGCCTGGAGACTGGCCGAGC 285
QY      785  TGGTGGGTCAGTGGCTCTTGTGGAGCGTGCCTGGGCTCCCAACCAAGCAGCTGCTGC 844
Db      284  TGGTGGGTCAGTGGCTCTTGTGGAGCGTGCCTGGGCTCCCAACCAAGCAGCTGCTGC 225
QY      845  CCGTGGAGATGCCCTGCATGAGAGCTGGCAAGTCTGTGGTGGCAAGTGAATTC 904
Db      224  CCGTGGAGATGCCCTGCATGAGAGCTGGCAAGTCTGTGGTGGCAAGTGAATTC 165
QY      905  CGAAGGACCACTTCTAAACAATGGACATGCTCACCTGGAAGTGGGTGGCCCAAGGCT 964
Db      164  CGAAGGACCACTTCTAAACAATGGACATGCTCACCTGGAAGTGGGTGGCCCAAGGCT 105
QY      965  ATCCCTCTGAGACATCTTAACTAGTGGGCAAGAGTCTCTGTCATCTGTCATCTGTAAGAGG 1024
Db      104  AGCCGCCGCAAGACVCTTAACTAGTGGGCAAGAGTCTCTGTCATCTGTCATCTGTAAGAGG 45
QY      1025  ATGACACC-ATCATGCAAGATTGGTAGATAATCATGG 1061
Db      44  ATGACACCTTNATGGGAGWATGGKDKRRTSDKG 7

RESULT 13
LOCUS      BUI87658
DEFINITION AGENCOURT_7851896 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6144738
5', mRNA sequence.
ACCESSION BUI87658
VERSION   BUI87658.1 GI:22701642
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13470 row: d column: 19
High quality sequence stop: 724.
FEATURES
    source
        1..889
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6144738"
            /tissue_type="retinoblastoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_67"
            /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.75 kb. Library constructed by Life
            Technologies."
ORIGIN
Query Match 76.2%; Score 823.4; DB 13; Length 889;
Best Local Similarity 99.1%; Pred. No. 1.9e-202;
Matches 838; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY      1  ATGGCGCTGGAGCTGGAGCTGTGTCGGGGGCGTGGTGGGGGCAACACCCGTGCTTC 60
Db      44  ATGGCGCTGGAGCTGGAGCTGTGTCGGGGGCGTGGTGGGGGCAACACCCGTGCTTC 103

```

```

QY      61  ATCATTTGCCAGATCGGCCAGAAACCAACAGGGGCGACCTTGGACCTAGCAAGCGCATGATC 120
Db      104  ATCATTTGCCAGATCGGCCAGAAACCAACAGGGGCGACCTTGGACCTAGCAAGCGCATGATC 163
QY      121  CGCATGGCCAAAGGCTGGGGCTGATTGTGCCAAGTCCAGAGAGAGTGGAGCTAGAAATTC 180
Db      164  CGCATGGCCAAAGGCTGGGGCTGATTGTGCTAAGTTCAGAGAGAGTGGAGCTAGAAATTC 223
QY      181  AAGTTTAAATCGGAAAGCCCTTGGAGGGCCATACACCTCGAAGCAATTCCTGGGGGAAAGCG 240
Db      224  AAGTTTAAATCGGAAAGCCCTTGGAGGGCCATACACCTCGAAGCAATTCCTGGGGGAAAGCG 283
QY      241  TAGGGGAGCACAACAGCATCTGGAGTTTACGCCATACAGTACAGGGAGCTGGAGAGG 300
Db      284  TAGGGGAGCACAACAGCATCTGGAGTTTACGCCATACAGTACAGGGAGCTGGAGAGG 343
QY      301  TAGCCCGAGGAGTGGGATCTTCTTCACTGCTCTGGCATGGATGGAGTGGAGTTGAA 360
Db      344  TAGCCCGAGGAGTGGGATCTTCTTCACTGCTCTGGCATGGATGGAGTGGAGTTGAA 403
QY      361  TTCCTGCATGAATGAATGTTCAATTTTCAAAGTTGGATCTGGAGACACTAAATAATTTT 420
Db      404  TTCCTGCATGAATGAATGTTCAATTTTCAAAGTTGGATCTGGAGACACTAAATAATTTT 463
QY      421  CCTTATCTGGAAGACAGCCCAAAAGGTGCCCAATGGTCTTCCAGTGGGATGCAG 480
Db      464  CCTTATCTGGAAGACAGCCCAAAAGGTGCCCAATGGTCTTCCAGTGGGATGCAG 523
QY      481  TCAATGGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 540
Db      524  TCAATGGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 583
QY      541  TTCTTGCAGTGPACAGGCGCATACCCGCTCCAGCTGAGGAGGAGTCAACCTCGGGTCATC 600
Db      584  TTCTTGCAGTGPACAGGCGCATACCCGCTCCAGCTGAGGAGGAGTCAACCTCGGGTCATC 643
QY      601  TCGGANTATCAGAAGCTCTTCTCAGCATTCCTCAGTGGTATCTTGGGATGAAACAGGC 660
Db      644  TCGGANTATCAGAAGCTCTTCTCAGCATTCCTCAGTGGTATCTTGGGATGAAACAGGC 703
QY      661  ATAGCGATATCTGTGGCGGAGTGGCTCTGGGGGGCAAGGTGTGGAAGCTCACATAACT 720
Db      704  ATAGCGATATCTGTGGCGGAGTGGCTCTGGGGGGCAAGGTGTGGAAGCTCACATAACT 763
QY      721  TTGGCAAGACCTGGAAGGGAGTGCACCTCGGCTTCCTGAGCTGAGGAGCTGGAGAGCTGGCC 780
Db      764  TTGGCAAGACCTGGAAGGGAGTGCACCTCGGCTTCCTGAGCTGAGGAGCTGGAGAGCTGGCC 823
QY      781  GAGCTGGTGGGCTCAGTGGC-TCCTTGTGAGGAGTGCCTTGGGCTCCCAACCAAGCAGCT 839
Db      824  GAGCTGGTGGGCTCAGTGGGCTCTTGTGGAGGAGTGGGCTCCCAACCAAGCAGCT 883
QY      840  GCTGCC 845
Db      884  GCTGCC 889

RESULT 14
LOCUS      BQ432284
DEFINITION AGENCOURT_7858941 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6167793
5', mRNA sequence.
ACCESSION BQ432284
VERSION   BQ432284.1 GI:21171360
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE

```


Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M13530 row: e column: 10
 High quality sequence stop: 711.
 Location/Qualifiers
 1..872
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6167793"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. primer: Oligo dt.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

JOURNAL
 COMMENT

FEATURES
 source
 1..872
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6167793"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. primer: Oligo dt.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

ORIGIN
 Query Match 74.4%; Score 803.2; DB 13; Length 872;
 Best Local Similarity 99.0%; Pred. No. 3.3e-197;
 Matches 808; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 1 ATCCGCTGAGCTGAGCTGTGCTCCGGGCGCTGGTGGGGGCAACACCGTGTCTC 60
 57 ATCCGCTGAGCTGGAGCTGTGCTCCGGGCGCTGGTGGGGGCAACACCGTGTCTC 116
 61 ATCATTTCCGAGATCGCCAGAACCCACGAGGAGCTGAGCTGAGCTGAGCTGATC 120
 117 ATCATTTCCGAGATCGCCAGAACCCACGAGGAGCTGAGCTGAGCTGAGCTGATC 176
 121 CGCATGCCAAGAGGTGTGGGCTGATTGTGCCAAGTTCAGAGAGAGTGTAGATTC 180
 177 CGCATGCCAAGAGGTGTGGGCTGATTGTGCCAAGTTCAGAGAGAGTGTAGATTC 236
 181 AAGTTTAATCGAAGCTTGGAGAGCCATACCTCGAAGCTTCTTGGGGAGAGC 240
 237 AAGTTTAATCGAAGCTTGGAGAGCCATACCTCGAAGCTTCTTGGGGAGAGC 296
 241 TAGCGGAGCACAACGACATCTGGAGTTCAGCCATGACAGTACAGGAGCTGCAGAG 300
 297 TAGCGGAGCACAACGACATCTGGAGTTCAGCCATGACAGTACAGGAGCTGCAGAG 356
 301 TAGCGGAGGAGTGTGGATCTTCTTCACTSCCTCTGCGCATGGATGAGATGCGCAT 360
 357 TAGCGGAGGAGTGTGGATCTTCTTCACTSCCTCTGCGCATGGATGAGATGCGCAT 416
 361 TTCTGCTGATCACTGAATGTTCATTTTCAAGTTGATCTGAGAGACACTAATAATTT 420
 417 TTCTGCTGATCACTGAATGTTCATTTTCAAGTTGATCTGAGAGACACTAATAATTT 476
 421 CCTTATCTGAAAGACAGACCAAAAGGTTCGCCAATGGTGTGATCTCCAGTGGATCAG 480
 477 CCTTATCTGAAAGACAGACCAAAAGGTTCGCCAATGGTGTGATCTCCAGTGGATCAG 536
 481 TCAATGACACCATGAAGAGCAAGTTTATCAGATCTGTGAAGCCCTTCAACCCCACTTCTGC 540
 537 TCAATGACACCATGAAGAGCAAGTTTATCAGATCTGTGAAGCCCTTCAACCCCACTTCTGC 596
 541 TTCTTGAGTGTACACCGGATACCCCGCTCAGCTGAGAGACGTCAACCTCGGGGTATC 600
 597 TTCTTGAGTGTACACCGGATACCCCGCTCAGCTGAGAGACGTCAACCTCGGGGTATC 656
 601 TCGAATATCAGAGCTCTTCTTCTGACATTTCCATAGGATTTCTGGGGCATGAACAGGC 660

Db 557 TCGGATATCAGAGCTCTTCTTCTGACATTTCCCATAGGATTTCTGGGCATGAACACG 716
 Qy 661 ATACGATATCTGTGGCGCAGTGGCTCTGTGGGGCAAGGTGTGGGAAGCTGACATAACT 720
 Db 717 ATACGATATCTGTGGCGCAGTGGCTCTGTGGGGCAAGGTGTGGGAAGCTGACATAACT 776
 Qy 721 TTGACACAGCTGGAAGGGAGTGAACACTTCGGCCTCGCTGGAGCTGGAGACTGGCC 780
 Db 777 TTGACACAGCTGGAAGGGAGTGAACACTTCGGCCTCGCTGGAGCTGGAGACTGGCC 836
 Qy 781 GAGCTGTGCGGTCAAGTGGCTCTTGTGGAGCGTGGC 816
 Db 837 GAGCTGTGCGGTCAAGTGGCTCTTGTGGAGCGTGGC 872

RESULT 15
 B0064521 1031 bp mRNA linear EST 02-42R-2002
 LOCUS AGENCOURT_6853542 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5926258
 DEFINITION 5', mRNA sequence.
 ACCESSION B0064521
 VERSION B0064521.1 GI:19893231
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1031)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapsb@mail.nih.gov
 Tissue Procurement: Lou Staudt
 cDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCV2099 row: e column: 11
 High quality sequence start: 22
 High quality sequence stop: 736.
 Location/Qualifiers
 1..1031
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5926258"
 /tissue_type="lymphoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_99"
 /note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the Laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."

ORIGIN
 Query Match 73.7%; Score 796.4; DB 13; Length 1031;
 Best Local Similarity 99.1%; Pred. No. 2.1e-195;
 Matches 811; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
 264 GGAGTTCAGGCATGACAGTACAGGAGCTGACAGGTACCGGAGGAGTGGATCTT 323
 Db 23 GGAGTTCAGGCATGACAGTACAGGAGCTGACAGGTACCGGAGGAGTGGATCTT 82
 Qy 324 CTTCACCTGCTCGCATGGATGAGATGSCAGTTGATTCCTCGCATGAATGATCTTC 383

Db 83 CTTCACTGCTCTGGGCTGGATGAGATGGAGTGTGAATCTCTCATGAACTGAATGTTCC 142
QY 384 ATTTTCAAAGTTGGATCTGGAGACACTAATAATTTTCCTTATCTGSAAGACAGCCAA 443
Db 143 ATTTTCAAAGTTGGATCTGGAGACACTAATAATTTTCCTTATCTGSAAGACAGCCAA 202
QY 444 ABAAGTCCGCCAATGGTGTCTCCAGTGGGATGCAAGTCAATGGACACCATGAAGCAAGT 503
Db 203 ABAAGTCCGCCAATGGTGTCTCCAGTGGGATGCAAGTCAATGGACACCATGAAGCAAGT 262
QY 504 TTATCAGATCGTGAAGCCCTCAACCCCACTTCTGCTTCTTCAGTGTACCAAGCCGATA 563
Db 263 TTATCAGATCGTGAAGCCCTCAACCCCACTTCTGCTTCTTCAGTGTACCAAGCCGATA 322
QY 564 CCGGCTCCAGCTGAGGAGCTCAACCTGCGGGTCATCTCGGAATATCAGAACTCTTTCC 623
Db 323 CCGGCTCCAGCTGAGGAGCTCAACCTGCGGGTCATCTCGGAATATCAGAACTCTTTCC 382
QY 624 TGACATTTCCCATAGGATATCTGGGCATGAACAGGCATAGCGATATCTGTGCCCGCAGT 683
Db 383 TGACATTTCCCATAGGATATCTGGGCATGAACAGGCATAGCGATATCTGTGCCCGCAGT 442
QY 684 GGCCTGGGGGCCAAGTGTGGAACTCACATACTTTGGACAAGACCTGGAAGGGGAG 743
Db 443 GGCCTGGGGGCCAAGTGTGGAACTCACATACTTTGGACAAGACCTGGAAGGGGAG 502
QY 744 TGACCACTCGGCTCGCTGGAGCTCGGAACCTGGCGAGCTGGTGGGTCACTGGTCT 803
Db 503 TGACCACTCGGCTCGCTGGAGCTCGGAACCTGGCGAGCTGGTGGGTCACTGGTCT 562
QY 804 TGTGGAGCTGCTGGGCTCCCAACCAAGCAGCTGCTGCCCTGTGAGATGGCCCTGCAA 863
Db 563 TGTGGAGCTGCTGGGCTCCCAACCAAGCAGCTGCTGCCCTGTGAGATGGCCCTGCAA 622
QY 864 TCAGAAGCTGGGCAAGTCTGTGGTGGCCAAAGTGAAATTCGGGAAGGCAACATTTAAC 923
Db 523 TCAGAAGCTGGGCAAGTCTGTGGTGGCCAAAGTGAAATTCGGGAAGGCAACATTTAAC 682
QY 924 AATGGACATGCTCAAGCTGGAAGTGGTGGAGCCCAAGCTATCTCTCTGAAGACATCTT 983
Db 683 AATGGACATGCTCAAGCTGGAAGTGGTGGAGCCCAAGCTATCTCTCTGAAGACATCTT 742
QY 984 TAATCTAGTGGGCAAGAGTCTGTGCTACTGTGTGAAGAGGATGACACCATCATGG-AA 1042
Db 743 TAATCTAGTGGGCAAGAGTCTGTGCTACTGTGTGAAGAGGATGACACCATCATGGAG 802
QY 1043 AATTGGTAGATTAATATGGCAAAATAATCAAGTCTTAA 1080
Db 803 AATTGGTAGATTAATATGGCAAAATAATCAAGTCTTAA 840

Search completed: May 11, 2004, 11:46:12
Job time : 3140.76 secs

361 TTCCTGCAATGAACTGATGTCATTTTCAAAGTTGGATCTGGAGACACTAATAATTTT 420
Db
361 TTCCTGCAATGAACTGATGTCATTTTCAAAGTTGGATCTGGAGACACTAATAATTTT 420
Qy
421 CTTTATCTGGAAAAGACAGCCAAAAGGTGCCCCAATGCTGATCTCCAGTGGGATGAG 480
Db
421 CTTTATCTGGAAAAGACAGCCAAAAGGTGCCCCAATGCTGATCTCCAGTGGGATGAG 480
Qy
481 TCAATGGACACCATGAAGCAAGTTTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 540
Db
481 TCAATGGACACCATGAAGCAAGTTTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 540
Qy
541 TTCCTGCAATGAACTGATGTCATTTTCAAAGTTGGATCTGGAGACACTAATAATTTT 600
Db
541 TTCCTGCAATGAACTGATGTCATTTTCAAAGTTGGATCTGGAGACACTAATAATTTT 600
Qy
601 TCGGAATATCAGAACTCTTTCTGATCATTCATAGGATTTCTGGGATGAACACAGGC 660
Db
601 TCGGAATATCAGAACTCTTTCTGATCATTCATAGGATTTCTGGGATGAACACAGGC 660
Qy
661 ATAGGATATCTGTCGCGCAGTGGCTCTGGGGGCCAAGTCTGGAAGTCACTAAT 720
Db
661 ATAGGATATCTGTCGCGCAGTGGCTCTGGGGGCCAAGTCTGGAAGTCACTAAT 720
Qy
721 TTGGACAAAGCTGGAAGGGAGTGACCACTCGGCTCGCTGGAGCTGGAGAACTGGGC 780
Db
721 TTGGACAAAGCTGGAAGGGAGTGACCACTCGGCTCGCTGGAGCTGGAGAACTGGGC 780
Qy
781 GAGCTGGTGGCTGATGCTGATGAGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 840
Db
781 GAGCTGGTGGCTGATGCTGATGAGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 840
Qy
841 CTGCTGCTGAGATGCTGATGAGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 900
Db
841 CTGCTGCTGAGATGCTGATGAGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 900
Qy
901 ATTCCGGAAGCACCATTCTAAATGAGACATGCTCAGCTGAGCTGGAGCTGGAGCTGGAG 960
Db
901 ATTCCGGAAGCACCATTCTAAATGAGACATGCTCAGCTGAGCTGGAGCTGGAGCTGGAG 960
Qy
961 GCCTATCTCTGAGACATCTTTTATCTAGTGGGAGAGAGCTCTGGTCTGATGTA 1020
Db
961 GCCTATCTCTGAGACATCTTTTATCTAGTGGGAGAGAGCTCTGGTCTGATGTA 1020
Qy
1021 GAGGATGACACCATCTGGAAGAAATGTTAGATAATCATGCAAAATAATCAAGTCTTA 1080
Db
1021 GAGGATGACACCATCTGGAAGAAATGTTAGATAATCATGCAAAATAATCAAGTCTTA 1080

RESULT 2

US-09-620-312D-1035
Sequence 1035, Application US/09620312D
Patent No. 6569662

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Exmanac, Radoje T.

TITLE OF INVENTION: No. 6569662el Nucleic Acids and

TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: PC FL-Genes Version 1.0
SEQ ID NO 1035
LENGTH: 1230
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (119)..(1198)
US-09-620-312D-1035

Query Match 99.7%; Score 1076.8; DB 4; Length 1230;
Best Local Similarity 99.8%; Pred. No. 1.1e-301;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCGCTGAGCTGGAGCTGTGCTCCCGGGGCGGTGGGTGGGGGGGCAACACCCGCTGCTC 60
Db 119 ATGCGCTGAGCTGGAGCTGTGCTCCCGGGGCGGTGGGTGGGGGGGCAACACCCGCTGCTC 178
Qy 61 ATCATTTGCCGATATGCGCCAGAACACACAGGGGAGCCTGGAGCTAGCCAGCGCATGATC 120
Db 179 ATCATTTGCCGATATGCGCCAGAACACACAGGGGAGCCTGGAGCTAGCCAGCGCATGATC 238
Qy 121 CCATGGCCCAAGAGTGTGGGCTGATTTCTGCCAGTTCACAGAGAGTGGAGTATGATTC 180
Db 239 CCATGGCCCAAGAGTGTGGGCTGATTTCTGCCAGTTCACAGAGAGTGGAGTATGATTC 298
Qy 181 AAGTTTAAATCGGAAAGCCTTGGAGAGCCATACACCTCGAAGCCTTCTGGGGGAAAGAG 240
Db 299 AAGTTTAAATCGGAAAGCCTTGGAGAGCCATACACCTCGAAGCCTTCTGGGGGAAAGAG 358
Qy 241 TACGGGGAGACAAAGAGCATCTGGAGTTCAGCCATGACAGTACAGGAGCTGAGAGG 300
Db 359 TACGGGGAGACAAAGAGCATCTGGAGTTCAGCCATGACAGTACAGGAGCTGAGAGG 418
Qy 301 TACGGGGAGAGGTTGGGATCTTCTCACTGCTCTGGCATGGATGAGATGGAGTTGAA 360
Db 419 TACGGGGAGAGGTTGGGATCTTCTCACTGCTCTGGCATGGATGAGATGGAGTTGAA 478
Qy 361 TTCCTGCAATGAACTGATGTCATTTTCAAAGTTGGATCTGGAGACACTAATAATTTT 420
Db 479 TTCCTGCAATGAACTGATGTCATTTTCAAAGTTGGATCTGGAGACACTAATAATTTT 538
Qy 421 CTTTATCTGGAAAAGACAGCCAAAAGGTGCCCCAATGCTGAGAGCCCTCAACCCCACTTCTGC 540
Db 539 CTTTATCTGGAAAAGACAGCCAAAAGGTGCCCCAATGCTGAGAGCCCTCAACCCCACTTCTGC 598
Qy 481 TCAATGGACACCATGAAGCAAGTTTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 540
Db 599 TCAATGGACACCATGAAGCAAGTTTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 658
Qy 541 TCTTGGAGTATCCAGGCGCATACCCGCTCCAGCTGAGGAGCTCAACCTCGGGGTATTC 600
Db 659 TCTTGGAGTATCCAGGCGCATACCCGCTCCAGCTGAGGAGCTCAACCTCGGGGTATTC 718
Qy 601 TCGGAATATCAGAACTCTTTCTGATCATTCATAGGATTTCTGGGATGAACACAGGC 660
Db 719 TCGGAATATCAGAACTCTTTCTGATCATTCATAGGATTTCTGGGATGAACACAGGC 778
Qy 661 ATAGGATATCTGTCGCGCAGTGGCTCTGGGGGCCAAGTCTGGAAGTCACTAAT 720
Db 779 ATAGGATATCTGTCGCGCAGTGGCTCTGGGGGCCAAGTCTGGAAGTCACTAAT 838
Qy 721 TTGGACAAAGCTGGAAGGGAGTGACCACTCGGCTCGCTGGAGCTGGAGAACTGGGC 780

Db 839 TTGACCAAGACTGGAGAGGGAGTGAACCACTCGGCTCGCTGGAGCTGGAGAACTGCC 898
QY 781 GAGCTGGTGGCTCAGTGCCTTGTGTGAGCTGCTGCCCTGGCTGCCCAACCAAGCAGCTG 840
Db 899 GAGCTGGTGGCTCAGTGCCTTGTGTGAGCTGCTGCCCTGGCTGCCCAACCAAGCAGCTG 958
QY 841 CTGCCCTGTGAGATGGCTGCAATGAGAGCTGGCAAGCTGTGGTGGCCCAAGTGA 900
Db 959 CTGCCCTGTGAGATGGCTGCAATGAGAGCTGGCAAGCTGTGGTGGCCCAAGTGA 1018
QY 901 ATTCGGAAGCCACCATCTTAACATCGACATGCTCACCGTGAGAGCTGGTGGCCCA 960
Db 1019 ATTCGGAAGCCACCATCTTAACATCGACATGCTCACCGTGAGAGCTGGTGGCCCA 1078
QY 961 GCCTATCTCTCTGAAGACATCTTAACTAGTGGCAAGAGCTCTGGTCACTGTGTGA 1020
Db 1079 GGCTATCTCTCTGAAGACATCTTAACTAGTGGCAAGAGCTCTGGTCACTGTGTGA 1138
QY 1021 GAGATGACACCATCATCGAAGAAATTTGTTAGATATCATGCGCAAAAATCAAGTCTTA 1080
Db 1139 GAGATGACACCATCATCGAAGAAATTTGTTAGATATCATGCGCAAAAATCAAGTCTTA 1138

RESULT 3
US-08-916-421B-1
; Sequence 1, Application US/38916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bull et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g

TOPOLOGY: linear
US-08-976-259-81

Query Match 4.9%; Score 52.6; DB 4; Length 1234;
Best Local Similarity 48.5%; Pred. No. 2.9e-05;
Matches 179; Conservative 0; Mismatches 184; Indels 6; Gaps 1;

QY 62 TCATTGCCAGATGGCCAGAACACACAGGGGACCTGAGCAGTACCCAGGCGCATGATCC 121
Db TGATGCCAGAGATTACTACTATCACTTTGGTGATATGAGACCGATTGAATCAATGATG 595
QY 122 GCATGGCCAGAGGTGGGGCTGATTTGCCAAGTCCAGAGAGTGGCTAGATTCATCA 181
Db TAGCGGCCACAGCGAGGGCTGACTATATCAAACTCGAGAGCGTGATGTTGAAGTT 655
QY 182 AGTTTAATCGAAAGCCCTTGAGAGGCCATACACCTCGAAGCATTCCTGGGGGAGAGCT 241
Db TCTATAGCAGGAGAGCTGGAGTCACCGTACCACTC-----TCTTTTGGCACCCT 709
QY 242 ACGGGGAGCACAAGCATCTGGAGTTACGCATGACAGTACAGGAGCTGCAGAGGT 301
Db TTAGGGACTATCGGCATGGCATTTGAATCAATGAAGAGCAATTTCTTTGTGCACTCT 769
QY 302 ACGCGGAGAGGTGGGATCTCTCACTGCTCTGGCATGATGAGATGGCAGTTGAAT 361
Db TCTGTAAGAGATTGGTATCGCTGGTTGCTTCTATTTAGATATGCCCTGCTAGAGT 829
QY 362 TCCTGCATGAATGAATTTCCATTTTCAAAAGTTGGATCTGGAGACACTAAATAATTC 421
Db TCATTGCGCAATTTGAACAGATATGATCAAGCTACCATCACTATATCTGAACATTAAG 889
QY 422 CTTATCTGG 430
Db 890 ATTATTTGG 898

RESULT 5

US-09-495-406-1
; Sequence 1, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarichuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/495,406
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; OTHER INFORMATION: including LOS biosynthesis locus
US-09-495-406-1

Query Match 4.1%; Score 44.4; DB 4; Length 11474;
Best Local Similarity 47.2%; Pred. No. 3.022;
Matches 169; Conservative 0; Mismatches 186; Indels 3; Gaps 1;

QY 363 CQTSATGAATCAATGCTTCCATTTTCAAAAGTTGGATCTGGAGACACTAATATTTCC 422
Db CTTAGAGATATGGGAGTTAGTCTTTTAAAGTTTCTAGATGGTTCAGGTGAGTGAATATATCC 7330
QY 423 TTATCTGGAAGACAGCGCCAAAGAGGTGCCCCAATGGTGATCTCCAGTGGGATCGAGTC 482
Db GCTTATTAACACATAGCAGCCCTTTAAAGCCCTATGATAGTAGTACACAGGGATGATAG 7390

QY 483 AATGGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGCTT 542
Db TATTGAAGATATAAAACCAACTGTAAATATCTTATTAGACATGAATTCCTTTTCTTTT 7450
QY 543 CTTGAGTGTACCGAGCGCATACCGCTCAGCTGAGGACGTCACCTCGGGTCACTC 602
Db AATGCACACCAACCAATCTTTTACCCCAACCCGATATCTTTTGAAGATTAAACGCTATGCT 7510
QY 603 GGAATATCAGAGCTCTTTTCTGACATTCCTGAGTATTCCTGGCATGAAACAGCAT 662
Db TGAATTAATAAAGAAATTTCTTTATGG---TAGCTTAAGCGACCAACACAGATTA 7567
QY 663 AGCGATATCTGTGGCGCAGTGGCTCTGGGGCCCAAGGTGTTTSSAACGTCACTAACT 720
Db TCTTCGGTGTAGTGGGTGCACTTTGGTCTTGTGCTTGAAGACATTTTACT 7625

RESULT 6

US-09-252-991A-6409
; Sequence 6409, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6409
; LENGTH: 1083
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6409

Query Match 4.0%; Score 43.4; DB 4; Length 1083;
Best Local Similarity 49.1%; Pred. No. 0.013;
Matches 144; Conservative 0; Mismatches 146; Indels 3; Gaps 1;

QY 26 CCGGCGCTGTGGTGGGGCAACACCCGCTTTCATCATTCGCGAGATCGGCCAGAAC 85
Db CCAAGCGGCTGTGTGTCATCGCCAGGGCGGAGTACATCAATCAACGCGTGCAGC 482
QY 86 ACNAGGGCGACCTGAGGTAGCCAGCGCATGATCCGATGCGCAAGAGGTGTGGGCTG 145
Db CCAATGGCTGTGTGTAGCGGTGACCCGCT---TACCGCTTCGAGCGAGCGTGGCTGG 539
QY 146 ATTGTGCCAAGTTCACAGAGGTGAGCTAGATTTCAAGTTTAAATCGGAAGCCCTGGAGA 205
Db ATGCGCCAGTTTCGCCAGCGCGCTTCGAGACCGATCACTGGCACTGGAAGAGG 599
QY 206 GGCCATACACTCGAAGCATTTCTGGGGGAGAGAGTACGGGGGACACAAACGACATCTGG 265
Db TCACCACCACTTGTGTCATCCGAGGGAAGCGTTCCGAGTGGTCAAGCTACCGACG 659
QY 266 AGTTCAGCCATGACAGGTAGCAGGAGCTGCGAGAGGTACGCGGAGGAGGTGGG 318
Db AACGCTGGGATGCCAGCTCAGCCCGCACTGCTGAACACTGTGTGTATGAGAG 712

RESULT 7

US-09-252-991A-6241
; Sequence 6241, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6241
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6241

```

Query Match	4.08; Score 43.4; DB 4; Length 1184;
Best Local Similarity	49.11; Ered. No. 0.013;
Matches 144; Conservative	0; Mismatches 146; Indels 3; Gaps 1;
QY	25 CC3GGGCGTGGGTGGGGCGGCAACACCGGTGCTTCATCATGCTCGAGATCGGCCAGAAC 85
Db	467 CCAAGCGGCTGTGTGGCATCGCCAGGGCGGAGTACATCCATCAACGCGGTGCAGC 526
QY	86 ACCAGGGGACCTCGAGTACCCAGCGCATGATCCGCATGGCCAGAGATGTGGGGCTG 145
Db	527 CCAATGGCTGTGTTACGGGTGACCCGC---TACCGCTTCGACGAGCACGGTGGCTCG 583
QY	146 ATTGTGCCAAGTTCGAGAAGGTGAGCTAGAATTCAAGTTTAATCGAAAGACCTTGGAGA 205
Db	584 AATCGGCCAGTTTTCGCCAGCGCGCGCTTCGAGACCGATCATCTGCCAACTGGAAGAGG 643
QY	206 GGCATACACCTCGAAGCATCTCTGGGGGAAGACGTTACGGGGAGCAAAACGACATCTGG 265
Db	644 TCACACCACTTGTGTGCATCGAGGGAAGCGTTCGGAGGTGGTCAAGCTACCGACCG 703
QY	266 AGTTCAGCATGACCAAGTACAGGAGCTGCAGAGGTACGCCGAGGAGGTTTGG 318
Db	704 AACGCTGGGATGCCAGCTCAGCCGCGCACTGCTGAACACTGTGGTGATGGAG 756

RESULT 8
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOMLOPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSES: Foley & Jardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768

```

; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match      3.8%; Score 41.4; DB 1; Length 7218;
Best Local Similarity 4.5%; Pred. No. 0.13;
Matches 18; Conservative 209; Mismatches 170; Indels 0; Gaps 0;

Qy      58  CAGGGCGACCTGGAGCTAGCCAGCGCATGATCCGATGCGCAGGAGTGTGGGCGTGAT 147
Db      1435  CRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1376
Qy      148  TGTGCCAAGTTCAGAAGACTGAGCTAGAATTCAAGTTTAATCGGAAGACCTTGGACAGG 207
Db      1375  RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1316
Qy      208  CCATACACCTCGAAGACATTCTCGGGGAGACACTACGGGAGACACAAACGACATCTGGAG 267
Db      1315  RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1256
Qy      268  TTCAGCATGACACGATCAGGAGGAGTGCAGAGTACGCCGAGGAGTGGGATCTTCTTC 327
Db      1255  RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1196
Qy      328  ACTGCCTCTGGCATGATGATGATGCGAGTGAATTCCTGCATGAACTGATTCATT 387
Db      1195  RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1136
Qy      388  TTCAAAGTTGGATCTCGGACACTAATAATTTTCCTATCTGAAAGACAGCAAAAAA 447
Db      1135  RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1076
Qy      448  GGTCCGCCAATGGTATCTCCAGTGGGATGCAAGTCAA 484
Db      1075  RRRRRRRRRATCGAAGCTCCCTCGACCTCGAGCAA 1039

```

```

RESULT 9
US-09-833-381-1478
: Sequence 1478, Application US/09833381
: Patent No. 6672186
: GENERAL INFORMATION:
: APPLICANT: Robison, Keith E.
: TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
: FILE REFERENCE: 5800-119
: CURRENT APPLICATION NUMBER: US/09/833,381
: CURRENT FILING DATE: 2001-04-11
: PRIOR APPLICATION NUMBER: 09/516,448
: PRIOR FILING DATE: 2000-02-29
: NUMBER OF SEQ ID NOS: 2050
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1478
: LENGTH: 2437
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: {1}...{2437}
: OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1478

```

Query Match 3.8%; Score 41; DB 4; Length 2437;

Best Local Similarity 52.7%; Pred. No. 0.095;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 724 GACAGACCTGGAAGGAGTGAACCTGCGCTGCTGGAGCTGGAGAACTGGCCGAG 783
Db 1803 GCCAAGCGGAGAGGACCGCAAGGCCATCGCCGCTTCCAGCAGGACAGCAGCGGCGG 1862
QY 784 CTGCTGGGTCAGTGGTCTTTGTGGAGCGTGCCTGGGCTCCCAACCAAGCAGCTGCTG 843
Db 1863 CAGGACAGCAGGAGGCGCCAGTGGGCGCCAGCCCGAGCGCCGAGCACTGCAGAGAAG 1922
QY 844 CCCTGTGAGATGCCCTCAATGAGAAGCTGGGCAAGCTGTGGTGCCA 892
Db 1923 CAGCCTGAGAAGGAGGAGGAGGAGCGCGGAGAGAGTGTGGAGACA 1971

RESULT 10

US-09-833-381-1479
; Sequence 1479, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1479
; LENGTH: 2437
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(2437)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1479

Query Match 3.8%; Score 41; DB 4; Length 2437;
Best Local Similarity 52.7%; Pred. No. 0.095;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 724 GACAGACCTGGAAGGAGTGAACCTGCGCTGCTGGAGCTGGAGAACTGGCCGAG 783
Db 1803 GCCAAGCGGAGAGGACCGCAAGGCCATCGCCGCTTCCAGCAGGACAGCAGCGGCGG 1862
QY 784 CTGCTGGGTCAGTGGTCTTTGTGGAGCGTGCCTGGGCTCCCAACCAAGCAGCTGCTG 843
Db 1863 CAGGACAGCAGGAGGCGCCAGTGGGCGCCAGCCCGAGCGCCGAGCACTGCAGAGAAG 1922
QY 844 CCCTGTGAGATGCCCTCAATGAGAAGCTGGGCAAGTCTGTGGTGCCA 892
Db 1923 CAGCCTGAGAAGGAGGAGGAGGAGCGCGGAGAGAGTGTGGAGACA 1971

RESULT 11

US-09-252-991A-13238
; Sequence 13238, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13238
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13238

Query Match 3.7%; Score 40; DB 4; Length 738;
Best Local Similarity 55.9%; Pred. No. 0.1;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 218 CGAAGCATCTCTGGGGAAGACGTACCGGGAGACAAACGACATCTGAGTTTCAGCCATG 277
Db 317 CGCCGCGTCTTGCAGAACATCTACAGGACTCAACCGGACCTGACATCCCGATTC 376
QY 278 ACCAGTACAGGAGCTGCAGAGGTACCGCGAGAGGTGGAGTCTTCTTCACTGCTCTG 337
Db 377 CCAACAACGGCTACCTGCAGCGCTGGCGGAGCAGGGCGTGTGTGCTCATACCTCC 436
QY 338 GCATGGATGAGATGGC 353
Db 437 TGACCGTCGAGCAGGC 452

RESULT 12

US-09-252-991A-9714
; Sequence 9714, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9714
; LENGTH: 363
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9714

Query Match 3.5%; Score 38.2; DB 4; Length 363;
Best Local Similarity 49.3%; Pred. No. 0.23;
Matches 100; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 697 AAGTGTGGAAGTGCATTAACCTTTGGACAAGACCTGGAGAGGAGTGAACCTCGGCC 756
Db 19 ATGATGCGCGACGCGCTCGCGCGCTGACCTGAGCGGTTGTGGAGCAACGACCGGCT 78
QY 757 TCGCTGGAGCTCGGAACTGGCGAGCTGGTGGGTCAGTGGCTCTTGTGGAGGTCGCC 816
Db 79 CCGCTGGAGCAGATGCGCTTGACCGAACAGCGCTGGAGCAGGCCAAGCGGTGGTGCC 138
QY 817 CTGGGCTCCCAACCAAGCAGCTGCTGCCCTGTGAGATGGCTGCAATGAGAAGCTGGC 876
Db 139 ACCGACGACGTGGCGGAGCTGAACTGGCCAGGACAAGTACGCGCGCGGCGAGATCGG 198
QY 877 AAGTGTGTTGGCGCAAGTGAA 899
Db 199 ATGACCGCGGAGTCTTACAGAA 221

RESULT 13

US-09-252-991A-9674
; Sequence 9674, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 9674

; LENGTH: 1254

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-9674

Query Match 3.5%; Score 38.2; DB 4; Length 1254;

Best Local Similarity 49.3%; Pred. No. 0.43;

Matches 100; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 697 AAGGTGTGGAACTGACATTAACCTTGGACAAGACCTGGAAGGGGAGTGACCACTCGGCC 756

Db 101 ATGATGGCCGAGGCTCGCGCGCTGACCTGAGCGTGTGTGAGCAACGACCGGCT 160

Qy 757 TCGTGGAGCTGGAGAACTGCGGAGCTGTGGGTCAGTSCGTCTTTGTGGAGCGTCCC 816

Db 161 CCGCTGGAGCAGATCGCGCTGACCAACAGCGCGCTGGAGCAGGCCAAGCGGTGGTCCC 220

Qy 817 CTGGGCTCCCAACCAAGCAGCTGCTGCCCTGTGAGATGGCTGCAATGAGAGCTGGGC 876

Db 221 ACCGACGAGTGGCGGAGCTGAACCTGGCCCGAGGACAAAGTACGCGCGCGATCGCG 280

Qy 877 AAGTCTGTGTGGCCAAAGTGAA 899

Db 281 ATGACCGCGAGTCTCTACAAGAA 303

RESULT 14

US-09-252-991A-9533/c

; Sequence 9533, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 9533

; LENGTH: 3654

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-9533

Query Match 3.5%; Score 38.2; DB 4; Length 3654;

Best Local Similarity 49.3%; Pred. No. 0.75;

Matches 100; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 697 AAGGTGTGGAACTGACATTAACCTTGGACAAGACCTGGAAGGGGAGTGACCACTCGGCC 756

Db 550 ATGATGGCCGAGGCTCGCGCGCTGACCTGAGCGTGTGCGAGCAACGACCGGCT 491

Qy 757 TCGTGGAGCTGGAGAACTGGCGAGCTGTGGGTCAGTSCGTCTTTGTGGAGCGTCCC 816

Db 490 CCGCTGGAGCAGATCGCGCTGACCAACAGCGCTGGAGCAGGCCAAGCGGTGGTCCC 431

Qy 817 CTGGGCTCCCAACCAAGCAGCTGCTGCCCTGTGAGATGGCTGCAATGAGAGCTGGGC 876

Db 430 ACCGACGAGTGGCGGAGCTGAACTGGCCCGAGGACAAAGTACGCGCGCGCAGATCGCG 371

Qy 877 AAGTCTGTGTGGCCAAAGTGAA 899

Db 370 ATGACCGCGAGTCTCTACAAGAA 348

RESULT 15

US-09-252-991A-13452

; Sequence 13452, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13452

; LENGTH: 486

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13452

Query Match

Best Local Similarity 3.4%; Score 36.4; DB 4; Length 486;

Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 196 GCCTTGGAGAGGCCATACACCTCGAAGCATTCCTGGGGGAGAGAGCTAGGGGAGCAGAA 255

Db 74 GCCTATGTGCGCCAGCGGAGTCTGTGCGGGGTGGCCCTTCGCGAGCAGACC 133

Qy 256 CGACATCTGGAGTTTCAGCCATGACCCAGTACAGGGGAGCTGCAGAGGTAGCCGAGGAGG 313

Db 134 CATCAGCAGCAGCGCGGAGGCGGAGGCGGAGCAGCAGCGGATCAGAGATAGGCGCGGAGGAG 191

Search completed: May 11, 2004, 07:22:56

Job time : 99.111 secs